GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:17:58 ; Search time 919.412 Seconds
(without alignments)
3343.258 Million cell updates/sec

Title: US-09-898-616A-1
Perfect score: 65
Sequence: 1 gatccatggaaatctgcccg.....tgctgatggacaccccgtcc 65

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	65	100.0	65	6	AR160915	AR160915 Sequence
c	2	42	64.6	42	6	AR160922	AR160922 Sequence
c	3	23	35.4	60	6	AR160921	AR160921 Sequence
	4	19	29.2	193659	10	AC122438	AC122438 Mus muscu
	5	19	29.2	215983	2	AC141530	AC141530 Rattus no
	6	19	29.2	220434	2	AC141151	AC141151 Rattus no
	7	18	27.7	39596	2	AL445257	AL445257 Homo sapi
	8	18	27.7	103513	9	AL357713	AL357713 Human DNA
c	9		27.7	110000	2	LMFLCHR36_11	Continuation (12 o
	10		27.7	156481	2	AC079838	AC079838 Homo sapi
			27.7	157650	9	AC007098	AC007098 Homo sapi
			27.7	167711	2	AC007629	AC007629 Homo sapi
			27.7	191854	1	AE017156	AE017156 Haemophil
			27.7	306790	8	AE016888	AE016888 Eremothec
			27.7	313050	1	BX321857	BX321857 Nitrosomo
			27.2	990	6	AX121458	AX121458 Sequence
			27.2	990	6	BD163575	BD163575 Novel pol
			27.2	1113	6	AX063919	AX063919 Sequence
			27.2	1113	6	AX244093	AX244093 Sequence
	20	17	26.2	3579	7	BPH298563	AJ298563 Bacteriop
	21	17	26.2	3592	7	BPH298557	AJ298557 Bacteriop
c	22	17	26.2	12111	1	AE004545	AE004545 Pseudomon
c	23	17	26.2	65674	5	BX936427	BX936427 Zebrafish
	24	17	26.2	81299	2	AC140044	AC140044 Mus muscu
c	25	17	26.2	110000	1	BX897699_10	Continuation (11 o
	26	17	26.2	149146	10	AC134848	AC134848 Mus muscu
	27	17	26.2	157321	9	AC068811	AC068811 Homo sapi
c	28	17	26.2	158219	9	AC005534	AC005534 Homo sapi
	29	17	26.2	163884	2	AC146388	AC146388 Pan trogl
c	30	17	26.2	164302	2	AC145691	AC145691 Mus muscu
	31	17	26.2	164423	2	CR361552	CR361552 Danio rer
c	32	17	26.2	168540	10	AC107673	AC107673 Mus muscu
c	33	17	26.2	169328	2	AC112659	AC112659 Mus muscu
c	34	17	26.2	171548	2	AC024732	AC024732 Homo sapi
c	35	17	26.2	175631	9	AC023511	AC023511 Homo sapi
c	36	17	26.2	177655	10	AC132288	AC132288 Mus muscu
c	37	17	26.2	179576	1	BX294156	BX294156 Pirellula
	38	17	26.2	181862	10	AC127262	AC127262 Mus muscu
c	39	17	26.2	182107	10	AC083915	AC083915 Mus muscu
c	40	17	26.2	185761	9	AC098974	AC098974 Homo sapi
c	41	17	26.2	186929	10	AC090657	AC090657 Mus muscu
	42	17	26.2	193627	2	AC091427	AC091427 Mus muscu
c	43	17	26.2	204632	2	AC097583	AC097583 Rattus no
	44	17	26.2	219405	2	AC118096	AC118096 Rattus no
c	45	17	26.2	227595	2	AC093396	AC093396 Bos tauru

ALIGNMENTS

RESULT 1
AR160915
LOCUS AR160915 65 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 6 from patent US 6255281.
ACCESSION AR160915
VERSION AR160915.1 GI:16225978
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 65)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
TITLE Use of recombinant human uteroglobin in treatment of inflammatory and fibrotic conditions
JOURNAL Patent: US 6255281-A 6 03-JUL-2001;
FEATURES Location/Qualifiers
source 1. .65
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 65; DB 6; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65
```

RESULT 2
AR160922/c
LOCUS AR160922 42 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 13 from patent US 6255281.
ACCESSION AR160922
VERSION AR160922.1 GI:16226000
KEYWORDS .
SOURCE Unknown.

```
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 13 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             64.6%;  Score 42;  DB 6;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-13;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 42
              |||||||||||||||||||||||||||||||||||||||||
Db         42 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 1


RESULT 3
AR160921/c
LOCUS       AR160921                  60 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 12 from patent US 6255281.
ACCESSION   AR160921
VERSION     AR160921.1  GI:16225996
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 12 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             35.4%;  Score 23;  DB 6;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 0.033;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 4
AC122438
LOCUS       AC122438              193659 bp    DNA     linear   ROD 05-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-216M6 from 5, complete sequence.
ACCESSION   AC122438
VERSION     AC122438.2  GI:22296783
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193659)
  AUTHORS   VanBrunt,A., Creason,K. and Van,Brunt.
  TITLE     The sequence of Mus musculus BAC clone RP24-216M6
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 193659)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 193659)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 193659)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 193659)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 17, 2002 this sequence version replaced gi:21105899.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: M_BB0216M06
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .193659
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /map="5"
                     /clone="RP24-216M6"
                     /clone_lib="RPCI-24"
     repeat_region   1978. .2053
                     /rpt_family="ID"
     misc_feature    2822. .3615
                     /note="CpG island (%GC=72.2, o/e=0.86, #CpGs=87)"
     repeat_region   4426. .4591
                     /rpt_family="L2"
     repeat_region   6139. .6322
                     /rpt_family="B2"
     tRNA            6144. .6216
                     /product="tRNA-Ser"
                     /note="Likely pseudogene (HMM Sc=33.90 / Sec struct
                     Sc=-11.21)"
     repeat_region   6385. .6621
                     /rpt_family="MER1_type"
```

```
repeat_region   6929. .7072
                /rpt_family="B4"
repeat_region   7131. .7183
                /rpt_family="ID"
repeat_region   7263. .7443
                /rpt_family="B2"
tRNA            complement(7366. .7438)
                /product="tRNA-Asn"
                /note="Likely pseudogene (HMM Sc=37.14 / Sec struct
                Sc=-11.21)"
repeat_region   7452. .7548
                /rpt_family="Alu"
repeat_region   7992. .8324
                /rpt_family="MaLR"
repeat_region   8343. .8501
                /rpt_family="L2"
repeat_region   8640. .8784
                /rpt_family="Alu"
repeat_region   8821. .8999
                /rpt_family="B2"
repeat_region   9048. .9094
                /rpt_family="Alu"
repeat_region   9298. .9522
                /rpt_family="B2"
repeat_region   9523. .9728
                /rpt_family="MaLR"
repeat_region   9729. .9842
                /rpt_family="Y-chromosomal"
repeat_region   9732. .9884
                /rpt_family="B4"
repeat_region   10020. .10059
                /rpt_family="B4"
repeat_region   10060. .10128
                /rpt_family="tRNA-Ala-GCY_"
repeat_region   10206. .10339
                /rpt_family="Alu"
repeat_region   10476. .10613
                /rpt_family="B4"
repeat_region   10665. .10864
                /rpt_family="B2"
repeat_region   11129. .11323
                /rpt_family="B2"
repeat_region   11878. .11944
                /rpt_family="ID"
repeat_region   12389. .12717
                /rpt_family="MaLR"
repeat_region   12879. .13035
                /rpt_family="B4"
repeat_region   13066. .13174
                /rpt_family="Alu"
repeat_region   14406. .14610
                /rpt_family="B2"
repeat_region   14626. .14705
                /rpt_family="ID"
repeat_region   15013. .15152
                /rpt_family="Alu"
repeat_region   15177. .15346
                /rpt_family="B2"
repeat_region   15438. .15709
                /rpt_family="L1"
repeat_region   15730. .15871
                /rpt_family="Alu"
repeat_region   16035. .16247
                /rpt_family="B2"
repeat_region   16262. .16487
                /rpt_family="B2"
repeat_region   16711. .16734
                /rpt_family="Alu"
repeat_region   16735. .16829
                /rpt_family="B4"
repeat_region   16936. .16999
                /rpt_family="ID"
repeat_region   17216. .17329
                /rpt_family="Alu"
repeat_region   17568. .17630
                /rpt_family="Alu"
repeat_region   17631. .17645
                /rpt_family="B4"
repeat_region   18027. .18228
                /rpt_family="B2"
repeat_region   18304. .18450
                /rpt_family="Alu"
repeat_region   18478. .18612
                /rpt_family="Alu"
repeat_region   18625. .18723
                /rpt_family="Alu"
repeat_region   18705. .18863
                /rpt_family="B2"
repeat_region   18864. .19010
                /rpt_family="Alu"
repeat_region   19011. .19024
                /rpt_family="B2"
repeat_region   19083. .19156
                /rpt_family="Alu"
repeat_region   19415. .19511
                /rpt_family="L2"
repeat_region   19747. .19926
                /rpt_family="B2"
repeat_region   20128. .20303
                /rpt_family="B2"
repeat_region   20542. .20647
                /rpt_family="Alu"
repeat_region   20653. .20793
                /rpt_family="B2"
repeat_region   20848. .20996
                /rpt_family="Alu"
repeat_region   21587. .21766
                /rpt_family="B2"
```

```
  Query Match             29.2%;  Score 19;  DB 10;  Length 193659;
  Best Local Similarity   100.0%;  Pred. No. 4.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          40 AAACCCTGCTGATGGACAC 58
               ||||||||||||||||||
Db       95628 AAACCCTGCTGATGGACAC 95646
```

```
RESULT 5
AC141530
LOCUS       AC141530               215983 bp    DNA     linear   HTG 17-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-517N11, WORKING DRAFT SEQUENCE, 47
            unordered pieces.
ACCESSION   AC141530
VERSION     AC141530.1  GI:28975794
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 215983)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
```

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 215983)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (17-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: KEFJ
Center clone name: CH230-517N11
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 200142 bases at least Q40
Consensus quality: 204905 bases at least Q30
Consensus quality: 208606 bases at least Q20
Estimated insert size: 204516; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1367: contig of 1367 bp in length
* 1368 1467: gap of unknown length
* 1468 2579: contig of 1112 bp in length
* 2580 2679: gap of unknown length
* 2680 4117: contig of 1438 bp in length
* 4118 4217: gap of unknown length
* 4218 5365: contig of 1148 bp in length
* 5366 5465: gap of unknown length
* 5466 6816: contig of 1351 bp in length
* 6817 6916: gap of unknown length
* 6917 7946: contig of 1030 bp in length
* 7947 8046: gap of unknown length
* 8047 9569: contig of 1523 bp in length
* 9570 9669: gap of unknown length
* 9670 11196: contig of 1527 bp in length
* 11197 11296: gap of unknown length
* 11297 12615: contig of 1319 bp in length
* 12616 12715: gap of unknown length
* 12716 14440: contig of 1725 bp in length
* 14441 14540: gap of unknown length
* 14541 15603: contig of 1063 bp in length
* 15604 15703: gap of unknown length
* 15704 17950: contig of 2247 bp in length
* 17951 18050: gap of unknown length
* 18051 20114: contig of 2064 bp in length
* 20115 20214: gap of unknown length
* 20215 21578: contig of 1364 bp in length
* 21579 21678: gap of unknown length
* 21679 24453: contig of 2775 bp in length
* 24454 24553: gap of unknown length
* 24554 26291: contig of 1738 bp in length
* 26292 26391: gap of unknown length
* 26392 27638: contig of 1247 bp in length
* 27639 27738: gap of unknown length
* 27739 30093: contig of 2355 bp in length
* 30094 30193: gap of unknown length
* 30194 33157: contig of 2964 bp in length
* 33158 33257: gap of unknown length
* 33258 36164: contig of 2907 bp in length
* 36165 36264: gap of unknown length
* 36265 39483: contig of 3219 bp in length
* 39484 39583: gap of unknown length
* 39584 42511: contig of 2928 bp in length
* 42512 42611: gap of unknown length
* 42612 46973: contig of 4362 bp in length
* 46974 47073: gap of unknown length
* 47074 50898: contig of 3825 bp in length
* 50899 50998: gap of unknown length
* 50999 54616: contig of 3618 bp in length
* 54617 54716: gap of unknown length
* 54717 58819: contig of 4103 bp in length
* 58820 58919: gap of unknown length
* 58920 64062: contig of 5143 bp in length
* 64063 64162: gap of unknown length
* 64163 67669: contig of 3507 bp in length
* 67670 67769: gap of unknown length
* 67770 72693: contig of 4924 bp in length
* 72694 72793: gap of unknown length
* 72794 76652: contig of 3859 bp in length
* 76653 76752: gap of unknown length
* 76753 83329: contig of 6577 bp in length
* 83330 83429: gap of unknown length
* 83430 88560: contig of 5131 bp in length
* 88561 88660: gap of unknown length
* 88661 94968: contig of 6308 bp in length
* 94969 95068: gap of unknown length
* 95069 100869: contig of 5801 bp in length
* 100870 100969: gap of unknown length
* 100970 108995: contig of 8026 bp in length
* 108996 109095: gap of unknown length
* 109096 116371: contig of 7276 bp in length
* 116372 116471: gap of unknown length
* 116472 121822: contig of 5351 bp in length
* 121823 121922: gap of unknown length
* 121923 127516: contig of 5594 bp in length
* 127517 127616: gap of unknown length

```
            *   127617   135246: contig of 7630 bp in length
            *   135247   135346: gap of unknown length
            *   135347   142259: contig of 6913 bp in length
            *   142260   142359: gap of unknown length
            *   142360   152944: contig of 10585 bp in length
            *   152945   153044: gap of unknown length
            *   153045   160462: contig of 7418 bp in length
            *   160463   160562: gap of unknown length
            *   160563   168267: contig of 7705 bp in length
            *   168268   168367: gap of unknown length
            *   168368   177703: contig of 9336 bp in length
            *   177704   177803: gap of unknown length
            *   177804   187092: contig of 9289 bp in length
            *   187093   187192: gap of unknown length
            *   187193   201446: contig of 14254 bp in length
            *   201447   201546: gap of unknown length
            *   201547   215983: contig of 14437 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215983
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-517N11"
ORIGIN

  Query Match             29.2%;  Score 19;  DB 2;  Length 215983;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          12 ATCTGCCCGTCTTTCCAGC 30
               ||||||||||||||||||||
Db      119509 ATCTGCCCGTCTTTCCAGC 119527


RESULT 6
AC141151
LOCUS       AC141151               220434 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-114H6, WORKING DRAFT SEQUENCE, 67
            unordered pieces.
ACCESSION   AC141151
VERSION     AC141151.2  GI:28951124
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220434)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220434)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220434)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 14, 2003 this sequence version replaced gi:28894506.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KEBT
                Center clone name: CH230-114H6
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 192918 bases at least Q40
                Consensus quality: 201911 bases at least Q30
                Consensus quality: 208432 bases at least Q20
                Estimated insert size: 198182; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 67 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1220: contig of 1220 bp in length
            *    1221     1320: gap of unknown length
            *    1321     2711: contig of 1391 bp in length
            *    2712     2811: gap of unknown length
            *    2812     4302: contig of 1491 bp in length
            *    4303     4402: gap of unknown length
            *    4403     5435: contig of 1033 bp in length
```

```
*      5436      5535: gap of unknown length
*      5536      6682: contig of 1147 bp in length
*      6683      6782: gap of unknown length
*      6783      8125: contig of 1343 bp in length
*      8126      8225: gap of unknown length
*      8226      9653: contig of 1428 bp in length
*      9654      9753: gap of unknown length
*      9754     10892: contig of 1139 bp in length
*     10893     10992: gap of unknown length
*     10993     12419: contig of 1427 bp in length
*     12420     12519: gap of unknown length
*     12520     13792: contig of 1273 bp in length
*     13793     13892: gap of unknown length
*     13893     15227: contig of 1335 bp in length
*     15228     15327: gap of unknown length
*     15328     16823: contig of 1496 bp in length
*     16824     16923: gap of unknown length
*     16924     18594: contig of 1671 bp in length
*     18595     18694: gap of unknown length
*     18695     20057: contig of 1363 bp in length
*     20058     20157: gap of unknown length
*     20158     21205: contig of 1048 bp in length
*     21206     21305: gap of unknown length
*     21306     22447: contig of 1142 bp in length
*     22448     22547: gap of unknown length
*     22548     24158: contig of 1611 bp in length
*     24159     24258: gap of unknown length
*     24259     25929: contig of 1671 bp in length
*     25930     26029: gap of unknown length
*     26030     27224: contig of 1195 bp in length
*     27225     27324: gap of unknown length
*     27325     28721: contig of 1397 bp in length
*     28722     28821: gap of unknown length
*     28822     29840: contig of 1019 bp in length
*     29841     29940: gap of unknown length
*     29941     31549: contig of 1609 bp in length
*     31550     31649: gap of unknown length
*     31650     33452: contig of 1803 bp in length
*     33453     33552: gap of unknown length
*     33553     35257: contig of 1705 bp in length
*     35258     35357: gap of unknown length
*     35358     37459: contig of 2102 bp in length
*     37460     37559: gap of unknown length
*     37560     38810: contig of 1251 bp in length
*     38811     38910: gap of unknown length
*     38911     39979: contig of 1069 bp in length
*     39980     40079: gap of unknown length
*     40080     41615: contig of 1536 bp in length
*     41616     41715: gap of unknown length
*     41716     44028: contig of 2313 bp in length
*     44029     44128: gap of unknown length
*     44129     45436: contig of 1308 bp in length
*     45437     45536: gap of unknown length
*     45537     48041: contig of 2505 bp in length
*     48042     48141: gap of unknown length
*     48142     50554: contig of 2413 bp in length
*     50555     50654: gap of unknown length
*     50655     52760: contig of 2106 bp in length
*     52761     52860: gap of unknown length
*     52861     54582: contig of 1722 bp in length
*     54583     54682: gap of unknown length
*     54683     57043: contig of 2361 bp in length
*     57044     57143: gap of unknown length
*     57144     59095: contig of 1952 bp in length
*     59096     59195: gap of unknown length
*     59196     62500: contig of 3305 bp in length
*     62501     62600: gap of unknown length
*     62601     64620: contig of 2020 bp in length
*     64621     64720: gap of unknown length
*     64721     67094: contig of 2374 bp in length
*     67095     67194: gap of unknown length
*     67195     68971: contig of 1777 bp in length
*     68972     69071: gap of unknown length
*     69072     71750: contig of 2679 bp in length
*     71751     71850: gap of unknown length
*     71851     74352: contig of 2502 bp in length
*     74353     74452: gap of unknown length
*     74453     77263: contig of 2811 bp in length
*     77264     77363: gap of unknown length
*     77364     80266: contig of 2903 bp in length
*     80267     80366: gap of unknown length
*     80367     83477: contig of 3111 bp in length
*     83478     83577: gap of unknown length
*     83578     86160: contig of 2583 bp in length
*     86161     86260: gap of unknown length
*     86261     88294: contig of 2034 bp in length
*     88295     88394: gap of unknown length
*     88395     93403: contig of 5009 bp in length
*     93404     93503: gap of unknown length
*     93504     98849: contig of 5346 bp in length

  Query Match             29.2%;  Score 19;  DB 2;  Length 220434;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         12 ATCTGCCCGTCTTTCCAGC 30
              ||||||||||||||||||
Db      26915 ATCTGCCCGTCTTTCCAGC 26933


RESULT 7
AL445257
LOCUS       AL445257               39596 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-1175N1, 14 unordered pieces.
ACCESSION   AL445257
VERSION     AL445257.1  GI:10716509
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: dJ1175N1
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 28716 bases at least Q40
            Consensus quality: 32676 bases at least Q30
            Consensus quality: 35450 bases at least Q20
            Insert size: 38296; sum-of-contigs
            Insert size: 138118; 12.9% error; agarose-fp
            Quality coverage: 1.51x in Q20 bases; sum-of-contigs Quality
            coverage: 0.97x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2942: contig of 2942 bp in length
            *     2943     3042: gap of 100 bp
```

```
            *     3043     5573: contig of 2531 bp in length
            *     5574     5673: gap of 100 bp
            *     5674     9075: contig of 3402 bp in length
            *     9076     9175: gap of 100 bp
            *     9176    11187: contig of 2012 bp in length
            *    11188    11287: gap of 100 bp
            *    11288    14656: contig of 3369 bp in length
            *    14657    14756: gap of 100 bp
            *    14757    17428: contig of 2672 bp in length
            *    17429    17528: gap of 100 bp
            *    17529    19964: contig of 2436 bp in length
            *    19965    20064: gap of 100 bp
            *    20065    24184: contig of 4120 bp in length
            *    24185    24284: gap of 100 bp
            *    24285    26468: contig of 2184 bp in length
            *    26469    26568: gap of 100 bp
            *    26569    29667: contig of 3099 bp in length
            *    29668    29767: gap of 100 bp
            *    29768    31926: contig of 2159 bp in length
            *    31927    32026: gap of 100 bp
            *    32027    34705: contig of 2679 bp in length
            *    34706    34805: gap of 100 bp
            *    34806    37009: contig of 2204 bp in length
            *    37010    37109: gap of 100 bp
            *    37110    39596: contig of 2487 bp in length.
FEATURES             Location/Qualifiers
     source          1. .39596
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-1175N1"
                     /clone_lib="RPCI-5"
     misc_feature    1. .2942
                     /note="assembly_fragment:00018
                     fragment_chain:1"
     misc_feature    3043. .5573
                     /note="assembly_fragment:00371
                     fragment_chain:1"
     misc_feature    5674. .9075
                     /note="assembly_fragment:00430
                     fragment_chain:2"
     misc_feature    9176. .11187
                     /note="assembly_fragment:00095
                     fragment_chain:2"
     misc_feature    11288. .14656
                     /note="assembly_fragment:00113"
     misc_feature    14757. .17428
                     /note="assembly_fragment:00120"
     misc_feature    17529. .19964
                     /note="assembly_fragment:00198"
     misc_feature    20065. .24184
                     /note="assembly_fragment:00245"
     misc_feature    24285. .26468
                     /note="assembly_fragment:00304"
     misc_feature    26569. .29667
                     /note="assembly_fragment:00328"
     misc_feature    29768. .31926
                     /note="assembly_fragment:00481"
     misc_feature    32027. .34705
                     /note="assembly_fragment:00489"
     misc_feature    34806. .37009
                     /note="assembly_fragment:00517"
     misc_feature    37110. .39596
                     /note="assembly_fragment:00532"
ORIGIN

  Query Match             27.7%;  Score 18;  DB 2;  Length 39596;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 ATCCATGGAAATCTGCCC 19
              ||||||||||||||||||
Db       9888 ATCCATGGAAATCTGCCC 9905


RESULT 8
AL357713
LOCUS       AL357713              103513 bp    DNA     linear   PRI 01-MAR-2001
DEFINITION  Human DNA sequence from clone RP11-3J23 on chromosome 1, complete
            sequence.
ACCESSION   AL357713 AC021953
VERSION     AL357713.11  GI:13234918
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103513)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On or before May 15, 2001 this sequence version replaced
            gi:7283183, gi:13160267.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-3J23 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-3J23 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-633I8 is at 103414 in this sequence.
            The true right end of clone RP5-1115A15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .103513
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-3J23"
                     /clone_lib="RPCI-11.1"
     repeat_region   870. .941
                     /note="36 copies 2 mer aa 65% conserved"
     repeat_region   877. .921
                     /note="5 copies 9 mer attaaataa 86% conserved"
     repeat_region   1178. .1469
                     /note="146 copies 2 mer aa 55% conserved"
     repeat_region   1477. .1752
                     /note="AluJo repeat: matches 3. .285 of consensus"
     repeat_region   2069. .2365
                     /note="AluSx repeat: matches 3. .300 of consensus"
     repeat_region   2599. .3429
```

```
                     /note="L1MC5 repeat: matches 6643. .7530 of consensus"
     repeat_region   4398. .4517
                     /note="MIR repeat: matches 23. .150 of consensus"
     repeat_region   5208. .5327
                     /note="L2 repeat: matches 2630. .2750 of consensus"
     repeat_region   6043. .6348
                     /note="AluSx repeat: matches 1. .306 of consensus"
     repeat_region   6359. .6657
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   6741. .6832
                     /note="L2 repeat: matches 2606. .2710 of consensus"
     repeat_region   6869. .7317
                     /note="LTR42 repeat: matches 7. .470 of consensus"
     repeat_region   7332. .7527
                     /note="L2 repeat: matches 2424. .2621 of consensus"
     repeat_region   9468. .9772
                     /note="AluJb repeat: matches 1. .299 of consensus"
     repeat_region   9806. .9927
                     /note="FLAM_C repeat: matches 6. .127 of consensus"
     repeat_region   10303. .10364
                     /note="MLT1J repeat: matches 1. .61 of consensus"
     repeat_region   10464. .10832
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   11305. .11465
                     /note="MIR repeat: matches 44. .197 of consensus"
     repeat_region   12077. .12376
                     /note="AluJo repeat: matches 1. .299 of consensus"
     repeat_region   12649. .12774
                     /note="63 copies 2 mer tt 65% conserved"
     repeat_region   12654. .12789
                     /note="8 copies 17 mer 64% conserved"
     repeat_region   12687. .12767
                     /note="9 copies 9 mer tttatttta 70% conserved"
     repeat_region   13172. .13399
                     /note="MER58A repeat: matches 1. .224 of consensus"
     repeat_region   13731. .14012
                     /note="AluJo repeat: matches 1. .277 of consensus"
     repeat_region   14292. .14592
                     /note="AluJb repeat: matches 1. .297 of consensus"
     repeat_region   17149. .17504
                     /note="THE1B repeat: matches 3. .364 of consensus"
     repeat_region   17952. .18217
                     /note="AluY repeat: matches 38. .303 of consensus"
     repeat_region   18219. .18518
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   18828. .18871
                     /note="11 copies 4 mer caca 100% conserved"
     repeat_region   18938. .18997
                     /note="30 copies 2 mer ac 96% conserved"
     repeat_region   19312. .19615
                     /note="AluSc repeat: matches 3. .303 of consensus"
     repeat_region   20942. .21167
                     /note="AluJb repeat: matches 72. .310 of consensus"
     repeat_region   21262. .21333
                     /note="MER5A repeat: matches 36. .104 of consensus"
     repeat_region   21479. .21559
                     /note="L1MC4 repeat: matches 7897. .7977 of consensus"
     repeat_region   21778. .21877
                     /note="25 copies 4 mer gaag 94% conserved"
     repeat_region   22337. .22632
                     /note="AluY repeat: matches 1. .299 of consensus"
     repeat_region   22662. .22972
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   23047. .23316
                     /note="AluJo repeat: matches 1. .299 of consensus"
     repeat_region   24280. .24579
                     /note="AluJb repeat: matches 1. .299 of consensus"
     repeat_region   24745. .24926
                     /note="AluYb8 repeat: matches 180. .308 of consensus"
     repeat_region   24927. .25226
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   25227. .25388
                     /note="AluYb8 repeat: matches 2. .180 of consensus"
     repeat_region   25427. .25721
                     /note="AluJo repeat: matches 1. .299 of consensus".
     repeat_region   26005. .26137
                     /note="AluSq/x repeat: matches 1. .136 of consensus"
     repeat_region   26138. .26443
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   27061. .27268
                     /note="AluJb repeat: matches 3. .209 of consensus"
     repeat_region   27510. .27820
                     /note="AluSg repeat: matches 1. .305 of consensus"
     repeat_region   27854. .28032
                     /note="MIR repeat: matches 85. .261 of consensus"
     repeat_region   28116. .28415
                     /note="AluY repeat: matches 3. .301 of consensus"
     repeat_region   28423. .28731
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   32979. .33020
                     /note="21 copies 2 mer ac 78% conserved"
     repeat_region   34272. .34561
                     /note="AluSq repeat: matches 9. .300 of consensus"
     repeat_region   35277. .35362
                     /note="L2 repeat: matches 2666. .2749 of consensus"
     repeat_region   35420. .35620
                     /note="MIR repeat: matches 16. .223 of consensus"
     repeat_region   37090. .37219
                     /note="L2 repeat: matches 2571. .2701 of consensus"
     repeat_region   38314. .38413
                     /note="MER94 repeat: matches 2. .100 of consensus"
     repeat_region   39688. .39711
                     /note="12 copies 2 mer aa 100% conserved"
     repeat_region   40371. .40654
                     /note="AluSx repeat: matches 6. .285 of consensus"
     repeat_region   42365. .42459
                     /note="AluJ/FLAM repeat: matches 1. .90 of consensus"
     repeat_region   42684. .42806
                     /note="FLAM_C repeat: matches 1. .123 of consensus"
     repeat_region   43618. .44082
                     /note="LTR19A repeat: matches 1. .486 of consensus"
     repeat_region   44567. .44813
                     /note="AluSx repeat: matches 44. .297 of consensus"
     repeat_region   47807. .47918
                     /note="AluJo/FRAM repeat: matches 169. .280 of consensus"
     repeat_region   47922. .47965
                     /note="11 copies 4 mer cata 86% conserved"
     repeat_region   48455. .48618
                     /note="FRAM repeat: matches 1. .160 of consensus"
     repeat_region   50339. .50623
                     /note="AluJo repeat: matches 1. .282 of consensus"
     repeat_region   50626. .50677
                     /note="26 copies 2 mer ac 84% conserved"
     repeat_region   50629. .50680
                     /note="13 copies 4 mer caca 84% conserved"
     repeat_region   50846. .51151
                     /note="AluJb repeat: matches 1. .306 of consensus"
     repeat_region   51947. .52235
                     /note="AluY repeat: matches 1. .289 of consensus"
     repeat_region   52994. .53120
                     /note="L2 repeat: matches 2565. .2703 of consensus"
     repeat_region   53182. .53488

  Query Match           27.7%;  Score 18;  DB 9;  Length 103513;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 ATCCATGGAAATCTGCCC 19
              ||||||||||||||||||
Db      80412 ATCCATGGAAATCTGCCC 80429


RESULT 9
LMFLCHR36_11/c
WPCOMMENT
  Sequence split into 36 fragments  LOCUS LMFLCHR36 Accession AL499624
```

Fragment Name	Begin	End
LMFLCHR36_00	1	110000
LMFLCHR36_01	100001	210000
LMFLCHR36_02	200001	310000
LMFLCHR36_03	300001	410000
LMFLCHR36_04	400001	510000
LMFLCHR36_05	500001	610000
LMFLCHR36_06	600001	710000
LMFLCHR36_07	700001	810000
LMFLCHR36_08	800001	910000
LMFLCHR36_09	900001	1010000
LMFLCHR36_10	1000001	1110000
LMFLCHR36_11	1100001	1210000
LMFLCHR36_12	1200001	1310000
LMFLCHR36_13	1300001	1410000
LMFLCHR36_14	1400001	1510000
LMFLCHR36_15	1500001	1610000
LMFLCHR36_16	1600001	1710000
LMFLCHR36_17	1700001	1810000
LMFLCHR36_18	1800001	1910000
LMFLCHR36_19	1900001	2010000
LMFLCHR36_20	2000001	2110000
LMFLCHR36_21	2100001	2210000
LMFLCHR36_22	2200001	2310000
LMFLCHR36_23	2300001	2410000
LMFLCHR36_24	2400001	2510000
LMFLCHR36_25	2500001	2610000
LMFLCHR36_26	2600001	2710000
LMFLCHR36_27	2700001	2810000
LMFLCHR36_28	2800001	2910000
LMFLCHR36_29	2900001	3010000
LMFLCHR36_30	3000001	3110000
LMFLCHR36_31	3100001	3210000
LMFLCHR36_32	3200001	3310000
LMFLCHR36_33	3300001	3410000
LMFLCHR36_34	3400001	3510000
LMFLCHR36_35	3500001	3529852

Continuation (12 of 36) of LMFLCHR36 from base 1100001 (AL499624 Leishmania major chromo

Query Match 27.7%; Score 18; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         14 CTGCCCGTCTTTCCAGCG 31
              ||||||||||||||||||
Db      66513 CTGCCCGTCTTTCCAGCG 66496
```

```
RESULT 10
AC079838
LOCUS       AC079838               156481 bp    DNA     linear   HTG 01-APR-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-257N14, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC079838
VERSION     AC079838.3  GI:13493151
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156481)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156481)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 1, 2001 this sequence version replaced gi:10799482.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0257N14
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 68%
            Sequencing vector: plasmid; 32%
            Chemistry: Dye-primer ET; 68% of reads
            Chemistry: Dye-terminator Big Dye; 32% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 154814 bases at least Q40
            Consensus quality: 155449 bases at least Q30
            Consensus quality: 155806 bases at least Q20
            Insert size: 139000; agarose-fp
            Insert size: 156381; sum-of-contigs
            Quality coverage: 6.40 in Q20 bases; agarose-fp
            Quality coverage: 6.28 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6178: contig of 6178 bp in length
            *     6179     6278: gap of unknown length
            *     6279   156481: contig of 150203 bp in length.
FEATURES            Location/Qualifiers
     source         1. .156481
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /clone="RP11-257N14"
     misc_feature   1. .6178
                    /note="assembly_name:Contig7
                    clone_end:T7
                    vector_side:right"
     misc_feature   6279. .156481
                    /note="assembly_name:Contig8
                    clone_end:SP6
                    vector_side:left"
ORIGIN
```

Query Match 27.7%; Score 18; DB 2; Length 156481;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         43 CCCTGCTGATGGACACCC 60
              ||||||||||||||||||
Db      11249 CCCTGCTGATGGACACCC 11266
```

```
RESULT 11
AC007098
LOCUS       AC007098               157650 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-443F16 from 2, complete sequence.
ACCESSION   AC007098
VERSION     AC007098.4  GI:11038585
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157650)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
```

REFERENCE 2 (bases 1 to 157650)
AUTHORS Kalicki,J., Johnson,D. and Harris,A.
TITLE The sequence of Homo sapiens BAC clone RP11-443F16
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 157650)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 157650)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 157650)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 30, 2000 this sequence version replaced gi:8570191.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0443F16
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-511I11, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-443F16; actual end is at base position 157456 of RP11-443F16.

FEATURES Location/Qualifiers
source 1. .157650
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-443F16"
/clone_lib="RPCI-11"

misc_feature 75. .449
/note="similar to EST H66784 (NID:g1025524) yr84h10.r1"
repeat_region 181. .425
/rpt_family="L1"
repeat_region 422. .1348
/rpt_family="L1"
repeat_region 1349. .1476
/rpt_family="Alu"
repeat_region 1477. .1870
/rpt_family="L1"
repeat_region 1871. .2176
/rpt_family="Alu"
repeat_region 2177. .2455
/rpt_family="L1"
repeat_region 2228. .2698
/rpt_family="L1"
repeat_region 2800. .3357
/rpt_family="MaLR"
repeat_region 3497. .3582
/rpt_family="L2"
repeat_region 3599. .3949
/rpt_family="L2"
repeat_region 4875. .4907
/rpt_family="L2"
repeat_region 5004. .5109
/rpt_family="L2"
repeat_region 5348. .5454
/rpt_family="L2"
repeat_region 6944. .7100
/rpt_family="MER2_type"
repeat_region 7099. .7878
/rpt_family="MER2_type"
repeat_region 7912. .8544
/rpt_family="MER2_type"
repeat_region 8545. .10505
/rpt_family="ERV1"
repeat_region 10510. .10614
/rpt_family="MER2_type"
repeat_region 10619. .10995
/rpt_family="MaLR"
repeat_region 11078. .11206
/rpt_family="L2"
misc_feature 12958. .13086
/note="similar to EST AA456682 (NID:g2179258) aa13e01.r1"
misc_feature 12958. .13086
/note="similar to EST AI150782 (NID:g3679251) qc06g03.x1"
misc_feature 12958. .13086
/note="similar to EST AI675083 (NID:g4875563) wc23c03.x1"
misc_feature 12965. .13086
/note="similar to EST AA721648 (NID:g2736631) ny87f02.s1"
misc_feature 13045. .13086
/note="similar to EST AW102800 (NID:g6073413) xd38c10.x1"
repeat_region 14003. .14303
/rpt_family="Alu"
repeat_region 14920. .15221
/rpt_family="Alu"
repeat_region 15257. .15564
/rpt_family="Alu"
repeat_region 16122. .16425
/rpt_family="Alu"
repeat_region 16796. .17077
/rpt_family="Alu"
repeat_region 17109. .17251
/rpt_family="Alu"
repeat_region 17636. .17941
/rpt_family="Alu"
repeat_region 17948. .18103
/rpt_family="Alu"
misc_feature 18524. .18766
/note="similar to EST Z21249 (NID:g27992)"
misc_feature 18526. .18766
/note="similar to EST AW102800 (NID:g6073413) xd38c10.x1"
misc_feature 18526. .18699

```
                     /note="similar to  EST AI675083 (NID:g4875563) wc23c03.x1"
     misc_feature    18526. .18691
                     /note="similar to  EST AA721648 (NID:g2736631) ny87f02.s1"
     misc_feature    18526. .18683
                     /note="similar to  EST AI150782 (NID:g3679251) qc06g03.x1"
     misc_feature    18526. .18606
                     /note="similar to  EST AA456682 (NID:g2179258) aa13e01.r1"
     misc_feature    18583. .18720
                     /note="similar to  EST AI806986 (NID:g5393552) wf24g10.x1"
     repeat_region   19222. .19264
                     /rpt_family="CR1"
     repeat_region   19386. .20704
                     /rpt_family="L1"
     repeat_region   20717. .20783
                     /rpt_family="L1"
     repeat_region   20832. .21073
                     /rpt_family="L1"
     repeat_region   21093. .21887
                     /rpt_family="L1"
     repeat_region   21888. .22187
                     /rpt_family="Alu"
     repeat_region   22188. .22460
                     /rpt_family="L1"
     repeat_region   22461. .22755
                     /rpt_family="Alu"
     repeat_region   22756. .23600
                     /rpt_family="L1"
     repeat_region   23981. .24220
                     /rpt_family="CR1"
     repeat_region   24228. .24530
                     /rpt_family="Alu"
     misc_feature    25833. .25891
                     /note="similar to  EST AA868764 (NID:g2964209) ak52f09.s1"
     repeat_region   26383. .26640
                     /rpt_family="L1"
     repeat_region   26840. .27084
                     /rpt_family="L1"
     repeat_region   27124. .27556
                     /rpt_family="L1"

  Query Match             27.7%;  Score 18;  DB 9;  Length 157650;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCC 60
               ||||||||||||||||||
Db       28647 CCCTGCTGATGGACACCC 28664


RESULT 12
AC007629/c
LOCUS       AC007629              167711 bp    DNA     linear   HTG 07-OCT-2000
DEFINITION  Homo sapiens clone NH0443B15a, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   AC007629
VERSION     AC007629.9  GI:10716584
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167711)
  AUTHORS   McCombie,W.R.
  TITLE     Human Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167711)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Oct 7, 2000 this sequence version replaced gi:9972279.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   142198: contig of 142198 bp in length
            *  142199   142298: gap of unknown length
            *  142299   157945: contig of 15647 bp in length
            *  157946   158045: gap of unknown length
            *  158046   160827: contig of 2782 bp in length
            *  160828   160927: gap of unknown length
            *  160928   163434: contig of 2507 bp in length
            *  163435   163534: gap of unknown length
            *  163535   165854: contig of 2320 bp in length
            *  165855   165954: gap of unknown length
            *  165955   167711: contig of 1757 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167711
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="NH0443B15a"
ORIGIN

  Query Match             27.7%;  Score 18;  DB 2;  Length 167711;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCC 60
               ||||||||||||||||||
Db      128804 CCCTGCTGATGGACACCC 128787


RESULT 13
AE017156
LOCUS       AE017156              191854 bp    DNA     linear   BCT 15-AUG-2003
DEFINITION  Haemophilus ducreyi strain 35000HP section 6 of 6 of the complete
            genome.
ACCESSION   AE017156 AE017143
VERSION     AE017156.1  GI:33149035
KEYWORDS    .
SOURCE      Haemophilus ducreyi 35000HP
  ORGANISM  Haemophilus ducreyi 35000HP
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE   1  (bases 1 to 191854)
  AUTHORS   Munson,R.S. Jr., Ray,W.C., Mahairas,G., Sabo,P., Mungur,R.,
            Johnson,L., Nguyen,D., Wang,J., Forst,C. and Hood,L.
  TITLE     The Complete Genome Sequence of Haemophilus ducreyi
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191854)
  AUTHORS   Munson,R.S. Jr., Ray,W.C., Mahairas,G., Sabo,P., Mungur,R.,
            Johnson,L., Nguyen,D., Wang,J., Forst,C. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2003) Pediatrics, Columbus Children's Research
            Institute and The Ohio State University, 700 Children's Drive,
            Columbus, OH 43205, USA
FEATURES             Location/Qualifiers
     source          1. .191854
                     /organism="Haemophilus ducreyi 35000HP"
                     /mol_type="genomic DNA"
                     /strain="35000HP"
                     /db_xref="taxon:233412"
     gene            complement(142. .2073)
                     /gene="thrS"
                     /locus_tag="HD1803"
     CDS             complement(142. .2073)
                     /gene="thrS"
                     /locus_tag="HD1803"
                     /codon_start=1
                     /transl_table=11
```

```
                /product="threonyl-tRNA synthetase"
                /protein_id="AAP96553.1"
                /db_xref="GI:33149036"
                /translation="MPIITLPDGSQRNFDNPVSVIEVAQSIGTGLAKATIAGRVNGER
                RDACDIIEQDSTLEIITAKDEDGLEIIRHSCAHLLGHAIKQLFPTVKMAIGPTIENGF
                YYDVDLDRTLTQDDLEAIEKRMLELAKTNYNVVKKPVSWQQARDTFEKRGEPYKIAIL
                DENIERTATPALYHHEEYIDMCRGPHVPNMRFCQHFKLQKVAGAYWRGDSKNKMLQRI
                YGTAWADKKQLTEYLTRLEEAAKRDHRRIGKALDLYHMQEEAPGMVFWHNDGWTIFRE
                LETPVRTKLKEYDYQEVKGPFMMDRVLWERTGHWQNYADLMFTTQSENREYAIKPMNC
                PGHVQIFNQGLKSYRDLPIRMAEFGSCHRNEPSGSLHGLMRVRGFTQDDAHIFCTEEQ
                IESEVTACIKMVYDIYSTFGFTNIFVKLSTRPDNRIGADDMWDRAEAGLSAALSANGL
                EYQIQHGEGAFYGPKIEPALRDCLGREWQCGTVQLDFALPSRLDASYVTEDNSRHTPV
                MIHRAILGSIERFIGIITEEYAGFFPVWLAPIQAVVMNITDSQADYVQQVVKQLSDAG
                LRVKADLRNEKVGFKVREHTLRRVPYMLVCGDKEITEGKVSVRTRKGEDLGTFNIEAF
                VEMLKAQVKTRELKLLGEA"
gene            2343. .3242
                /gene="pnnK"
                /locus_tag="HD1804"
CDS             2343. .3242
                /gene="pnnK"
                /locus_tag="HD1804"
                /codon_start=1
                /transl_table=11
                /product="probable inorganic polyphosphate/ATP-NAD kinase"
                /protein_id="AAP96554.1"
                /db_xref="GI:33149037"
                /translation="MNTMKNVTKRSFQTIAIVGKPRHDNALETHLAVYNWLKDRHYSV
                LVEEKIAEQLQLPNGKRIEEIGQIADLVIVIGGDGNMLGMARSLAKYQVPLIGINRGN
                LGFLTDIAPQSAFEQLYSCLEKGEFIIEQRFLLEAQIEQNGKIISANNALNEVAIHPT
                QVARIIEFEVYIDSKFAFSQRSDGLIIATPTGSTAYSLSAGGPILTPNMNAIALVPMH
                PHALSSRPLVIDGDSHISLRFAQYNQPNLEISCDGQDDLPPTPEDRIIVRKSPDILHL
                LHLKDYNYFTVLGSKLGWSSKLF"
gene            complement(3274. .3867)
                /locus_tag="HD1805"
CDS             complement(3274. .3867)
                /locus_tag="HD1805"
                /codon_start=1
                /transl_table=11
                /product="putative MAF protein"
                /protein_id="AAP96555.1"
                /db_xref="GI:33149038"
                /translation="MHKKLYLASNSPRRWALLQNLGLALLPLASEIDETAHANEMAQD
                YCSRIAREKNQAAQAVRIRQNLAEYPILTADIMVSIDGEILGKPSHQQQAVEMLKKLS
                GRTHQVYTAVCVSANQQLFECIHSSEVTFKRLSEAEIYAYIATGEPMDKAGAYGIQAL
                GGIFIQHLAGSFTGVMGLPIFETVALLKKVNIEVLPS"
gene            complement(3940. .5394)
                /gene="thiI"
                /locus_tag="HD1806"
CDS             complement(3940. .5394)
                /gene="thiI"
                /locus_tag="HD1806"
                /codon_start=1
                /transl_table=11
                /product="thiamine biosynthesis protein ThiI"
                /protein_id="AAP96556.1"
                /db_xref="GI:33149039"
                /translation="MKFIIKLFPEIMIKSDSVRKRFIKILTSNIRNVLLRETEQVAVV
                RNWDFIEVRAKVAEEIPLILDLLKRTPGIHHILEVQEMPFTSLHDIFEHTFAKYQNQL
                VDKTFCVRVRRKGKHEFSSLDVEKYVGGGLNQRIESARVKLTKPDVTVRIDINGDKML
                LVEARHEGLGGYPIGTQEDVLSLISGGFDSGVSSYMFIRRGSRVHYCFFNLGGASHEI
                GVKQMAYHIWSRYSTSHKVRFVAINFESVVGEILEKVDNGQMGVVLKRMMVRAASQIA
                ERFAIQAIVTGEALGQVSSQTLTNLRLIDKAADSLVLRPLITHDKEKIIALAKQIGTD
                DIAKSMPEFCGVISKNPTVKAIESKIVEEEGHFDFDVLEKAVQNATYLDIREIALQTE
                KDVVAVEATSALTEKDIILDIRSPEEMDEKPLVLAQAQVIELPFYKLSTQFAHLDQSK
                NYLLYCERGVMSKLQALYLKEKGYQNVKVFNLPK"
gene            complement(5519. .6817)
                /gene="purA"
                /locus_tag="HD1807"
CDS             complement(5519. .6817)
                /gene="purA"
                /locus_tag="HD1807"
                /codon_start=1
                /transl_table=11
                /product="adenylosuccinate synthetase"
                /protein_id="AAP96557.1"
                /db_xref="GI:33149040"
                /translation="MGKSVAILGAQWGDEGKGKIVDLLTDRVKYVVRYQGGHNAGHTL
                IINGEKTVLRLIPSGILRENVTCLIGNGVVLSPEALMKEMGELEACGINVRERLKISE
                ACPLILPYHVAMDHAREAALGKNKIGTTGRGIGPAYEDKVARRGLRVSDLFDKEAFAT
                KLKDILDLYNFQLVHYYKVEPVDFQKTLDDVFAVADVIKGMVADVTTLLHQARKEGVN
                ILFEGAQGTMLDIDHGTYPFVTSSNTTAGGVATGAGFGPRNLDYVLGIIKAYCTRVGS
                GPFTTELFNEEGETIARKGNEFGAVTGRPRRCGWFDAVAVRRAVQINSISGFCMTKLD
                VLDGFETLKICTAYKMPNGEIVEYAPMAAKDWEGVEPIYETMPGWSENTFGVIKYEAL
                PQAALDYIKRIEELVGVPVDILSTGPDRIETMILRDPFVA"
gene            complement(6953. .7840)
                /gene="hflC"
                /locus_tag="HD1808"
CDS             complement(6953. .7840)
                /gene="hflC"
                /locus_tag="HD1808"
                /codon_start=1
                /transl_table=11
                /product="HflC protein"
                /protein_id="AAP96558.1"
                /db_xref="GI:33149041"
                /translation="MRKLLLPIVSLVMMALISCLVIVPEGYRGIMLRFNKVQRDADQK
                VVVYEPGLHVKVPFIDSLKILDSRIQMLDDQEDRFVTVEKKDLLVDSYVKWRISDFGK
                FYTATGGDVKRASDLLRRKVGDRLRSEIGSRTIKDIVSGSRGELMAGAQKALNAGEDG
                AEKLGIEVVDVRVKQINLPKEVSSSIYQRMRAERDAVAREHRSQGEEKAEFIRAEVDK
                KVILIEEANAKKKAEILRGEGDAIAAKIYAEAFSKAPDFYSFVRSLKAYENSFTKDQQN
                MMLLKSDSEFFRFMKAPTK"
gene            complement(7843. .9048)
                /gene="hflK"
                /locus_tag="HD1809"
CDS             complement(7843. .9048)
                /gene="hflK"
                /locus_tag="HD1809"
                /codon_start=1
                /transl_table=11
                /product="HflK protein"
                /protein_id="AAP96559.1"
                /db_xref="GI:33149042"
                /translation="MSWNESGNQQDPWGKPGQKKPEQQEPSGQGNQQDPNRPKSAQNP
                PDLEEAFGKLLKKMGGGNKHSNNQRPASLGKFIPAIVIFSALVWGASGFYTVQEAERG
                VVTRFGKLHQIVMPGLNWKPTFIDQVIPINIERVSELKTQGSMLTQDENMVQVEMTVQ
                YRVEDPAKYKFSVRNADDSLKQATDSALRYVIGHMSMDDILTKGRATVREKTWETLRE
                IIKTYDMGLLVTDVNFQSARPPEEVKDAFDDAIKAQEDEQRLIREAEAYARGREPLAR
                GQAQRIIEQATAYKEQIVLEAQGDIQRFSKLLPEYQAAPAVMRERLYIETMEKVMKNT
                PKIIMDSNSNNVNVLPLEKFLGKTTASEVAQKPLATPVQQPQVSHQLAPSVVTEDRNR
                EQPIRRGRF"
gene            complement(9247. .9564)
                /gene="trxA"
                /locus_tag="HD1810"
CDS             complement(9247. .9564)
                /gene="trxA"
                /locus_tag="HD1810"
                /codon_start=1
                /transl_table=11
                /product="thioredoxin"
                /protein_id="AAP96560.1"
                /db_xref="GI:33149043"
                /translation="MTHQVTDATFEQEVLKSDLPVLLDFWAPWCGPCRTIAPWLDELA
                QEFAGRAKVAKVNVDENQQIAAQFGIRSIPTLLLFKNGEVVAIQVGVLPKSQLVTFIE
                QAL"
gene            complement(9663. .10955)
                /locus_tag="HD1811"
CDS             complement(9663. .10955)
                /locus_tag="HD1811"
                /codon_start=1
                /transl_table=11
                /product="putative xanthine permease"
                /protein_id="AAP96561.1"
                /db_xref="GI:33149044"
                /translation="MQNLRYPIDSKPPFGLSVLLAVQHLLAALGGIIAVPLVIGNVLK
                LPTSDTIVLVNAALLVSGIVTIIQCRGIGPIGLRLPSVMGTSFTFVAAALAIGFSEYG
                VAGIMGSSLIGSLVMIIGSFFMPYVRKLFPPVVTGIVVMMIGLSLIPVAVDWFAGGQK
```

```
                     GDPHYADPANLAMATFVLVLVVILVQWGKGIFSAAAIVIGMLVGYAVALALGWINFEA
                     IQNAEAFAIPQPLHFGLAFPISGIIGMSIAYLVTIVESSGNFLALGNATQTEITGKHL
                     RGGVLCDGLGSAFAAVMSTTPFSSFAQNIGVISLTGVASRYVVTVMGVLLVLAGVFPW
                     LGAIIVSIPSPVLGGAGLMMFAMIIAAGIQMLDKVERSKRNGLIIAISIGCGLAVTTR
                     PELLDKLPNFFKEVFGSGITVGSVLALLLNLILPEDKH"
     gene            complement(11033. .12871)
                     /gene="uvrC"
                     /locus_tag="HD1812"
     CDS             complement(11033. .12871)
                     /gene="uvrC"
                     /locus_tag="HD1812"
                     /codon_start=1
                     /transl_table=11
                     /product="excinuclease ABC, subunit C"
                     /protein_id="AAP96562.1"
                     /db_xref="GI:33149045"
                     /translation="MSQFNAKAFLADVPHLPGVYRMYDANQAIIYVGKAKDLKKRLAS
                     YFRSQVASKKIEALVGNIHHIDTTITHSETEALLLEHNYIKQHQPKYNVLLRDDKSYP

  Query Match             27.7%;  Score 18;  DB 1;  Length 191854;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         40 AAACCCTGCTGATGGACA 57
              ||||||||||||||||||
Db     119863 AAACCCTGCTGATGGACA 119880


RESULT 14
AE016888
LOCUS       AE016888               306790 bp    DNA     linear   PLN 14-APR-2004
DEFINITION  Eremothecium gossypii chromosome III, section 3 of 3 of the
            complete sequence.
ACCESSION   AE016888 AE016816
VERSION     AE016888.1  GI:44981618
KEYWORDS    .
SOURCE      Eremothecium gossypii (Ashbya gossypii)
  ORGANISM  Eremothecium gossypii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Eremothecium.
REFERENCE   1  (bases 1 to 306790)
  AUTHORS   Dietrich,F.S., Voegeli,S., Brachat,S., Lerch,A., Gates,K.,
            Steiner,S., Mohr,C., Pohlmann,R., Luedi,P., Choi,S., Wing,R.A.,
            Flavier,A., Gaffney,T.D. and Philippsen,P.
  TITLE     The Ashbya gossypii genome as a tool for mapping the ancient
            Saccharomyces cerevisiae genome
  JOURNAL   Science 304 (5668), 304-307 (2004)
   PUBMED   15001715
REFERENCE   2  (bases 1 to 306790)
  AUTHORS   Brachat,S., Voegeli,S.E., Dietrich,F.S., Lerch,A., Gaffney,T. and
            Philippsen,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2002) Applied Microbiology, Biozentrum,
            Universitaet Basel, Klingelbergstrasse 50-70, Basel CH-4056,
            Switzerland
COMMENT     This is low coverage sequence generated to identify the complete
            set of genes and the gene order on this chromosome. Regions of low
            quality are identified. Before doing extensive work on any gene
            identified here the sequence should be confirmed.
FEATURES             Location/Qualifiers
     source          1. .306790
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /strain="ATCC 10895"
                     /db_xref="ATCC:10895"
                     /db_xref="taxon:33169"
                     /chromosome="III"
                     /note="synonym: Ashbya gossypii"
     source          1. .44005
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG1132"
     source          38756. .123538
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG1591"
     source          115294. .185251
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG1038"
     source          160613. .238543
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG1820"
     source          213860. .273940
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG2333"
     source          232620. .295242
                     /organism="Eremothecium gossypii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:33169"
                     /clone="BAC clone bAG2449"
     unsure          692
                     /note="region of low quality - less than 90 percent
                     confidence"
     unsure          706. .707
                     /note="region of low quality - less than 90 percent
                     confidence"
     gene            <1102. .>2283
                     /locus_tag="ACR143W"
     mRNA            <1102. .>2283
                     /locus_tag="ACR143W"
                     /product="ACR143Wp"
     CDS             1102. .2283
                     /locus_tag="ACR143W"
                     /note="syntenic homolog of Saccharomyces cerevisiae
                     YPL154C (PEP4); tandem gene duplication in Ashbya
                     gossypii"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ACR143Wp"
                     /protein_id="AAS51369.1"
                     /db_xref="GI:44981619"
                     /translation="MKLQLLSLLVLLVPSISAENGVFYAQVYRDGVDVVSNPELISRH
                     VEAVAARHELSRRTVEDEELQVGGGKFSANVYNFENFQYSVDITLGTPAQNFRVALDT
                     GSSLLWIPSDRCTSQICRTRNRYRSGASSTFKATDKTLRLQYVKGDAKARVSYDTLYF
                     AGAKIENQGFGEAEAIGDDFSGARFDGIIGIGYPSIGYGIKPPINTLIDSGGLKDPMF
                     GIYISNAQNRNSPVGEIVLGGYNTQKFKGDIKWLPVLRKAFWETDLSAFKVGNFALDV
                     QGLTAVFDTGSSFIIMPEDTYTDPVSQFPGASHDDGTDYVDCSTVNSGPDLTLDFGGV
                     ALKLSARDYVMQLGRNTCILAVSGNSRVTGEVILGDTFLRRYYTIYNFGDNTIGVASA
                     V"
     gene            <3039. .>4265
                     /locus_tag="ACR144W"
     mRNA            <3039. .>4265
                     /locus_tag="ACR144W"
                     /product="ACR144Wp"
     CDS             3039. .4265
                     /locus_tag="ACR144W"
                     /note="syntenic homolog of Saccharomyces cerevisiae
                     YPL154C (PEP4); tandem gene duplication in Ashbya
                     gossypii"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="ACR144Wp"
                     /protein_id="AAS51370.1"
                     /db_xref="GI:44981620"
                     /translation="MRFQTLLPLAALLAPSFANAEGGVYKATLHKQSFDITPEELSMK
                     QRAASVGQKYLNALSKVVSDPSIQNEYNLFAKNNDGHVSELANLANNIYAADVTIGTP
                     PQDFRVVVDTGSSTLWVPGKECRAMACRLHKRYDHDRSSTYKENGTLTGVTYGSGSIM
                     GYVSEDTFRISDLEIPGQQFTETTDEPGSVFVFAAFDGILGLAYPSLGYGVTPPFQQL
```

```
                MEKKLVKEPVFGMYLDDMKTGEGNGQLVLGGYDETKFKGEIAWLPVRRQAYWEVVFDS
                VTVGDDNIPLENYGAAIDSGTSLITFPSELFNQFISKLSGVTKDRQGNVYVDCANKQT
                APPLTFGFGGHKFSISGEDYIISVPGQSARCMPAIVQLDIDSAGKVAIIGDVFLRRYY
                SIYDFGNNAVGLATAV"
gene            <4847. .>6904
                /locus_tag="ACR145W"
mRNA            <4847. .>6904
                /locus_tag="ACR145W"
                /product="ACR145Wp"
CDS             4847. .6904
                /locus_tag="ACR145W"
                /note="syntenic homolog of Saccharomyces cerevisiae
                YPL155C (KIP2)"
                /codon_start=1
                /evidence=not_experimental
                /product="ACR145Wp"
                /protein_id="AAS51371.1"
                /db_xref="GI:44981621"
                /translation="MMMVARVATAEHVGSAGASLPQTPGSRSFALGAHPGPQKRIGGP
                AQGQTAFITPLVTPDGIYSRPSSPYMQASPLLKGSESGGSAGSPQSPDAPSSASGASV
                GNAIGSGYTGNVSVAIRIKPSESSTKDPWYASNDRLIHTEFGEFQFDHVFTKGVCNQE
                VYQALGVPIIDKLFEGYNATIFAYGMTGSGKTFTMSGNKQEPGLIPQCVGNIFDRISS
                EHHGASLAYEVKVSYLEIYNEKIYDLLNYVDRQAGSTGQPSRNATGLKIRDDSKYGVK
                VVDLTEQLVSSHEDVMKWIATGDRNRKTGETDFNTRSSRSHAIVLLRLTRYDLKTGSE
                ATSTLSLCDLAGSERAVTQIVRRKEGAFINKSLLALGTVIAKLSMLGSQANGLQPSPA
                AGHIPYRDSKLTRILQPALTGDSIITTICTIDSKAESSTETTNTVRFASRAKNIALNV
                RKNEMDSHAEKDTIIQNLRKQLDEQHETIVMLRRSAAAPSGNGSTSPLDSPGVGGTSL
                SERTHNMEKGLLEVENSILKTKLEHCEKLLDKDMMVLEDPHVREIVEMLPLDIASVLE
                SKVQGMESQLRQYRVVYVQKLESDLLKAQRNIITTHSVQFNRQSTANVQEKYGSDVDIE
                LLLEEQEAELMELRNALKRKDKMIEALQSARRLRDSALSPTTTVLLQRKESVIDPRDP
                QPV"
gene            <7458. .>8099
                /locus_tag="ACR146W"
mRNA            <7458. .>8099
                /locus_tag="ACR146W"
                /product="ACR146Wp"
CDS             7458. .8099
                /locus_tag="ACR146W"
                /note="syntenic homolog of Saccharomyces cerevisiae
                YPL156C (PRM4)"
                /codon_start=1
                /evidence=not_experimental
                /product="ACR146Wp"
                /protein_id="AAS51372.1"
                /db_xref="GI:44981622"
                /translation="MKPKVQSRKGFRALWMAATLLTVAVYLCVAHNRGWYDSPQVRTP
                QRIPTSSGAAPEVWEPAKPRATRPLVNLDAVRRGGGDFRPAQNFQEIISTSPFVLFVK
                GTDESSRYLRHLLTNEYEVTPEIAVVDLSKHRYSGDLQRYIREHKLHGPLFQLTENTP
                DVPYLFVKGESVINTSVEKDIREPHERGQLQAELQRHAGLDAVIKKRSPPSNS"
gene            complement(<8145. .>9080)
                /locus_tag="ACR147C"
mRNA            complement(<8145. .>9080)
                /locus_tag="ACR147C"
                /product="ACR147Cp"
CDS             complement(8145. .9080)
                /locus_tag="ACR147C"
                /note="syntenic homolog of Saccharomyces cerevisiae
                YPL157W"
                /codon_start=1
                /evidence=not_experimental
                /product="ACR147Cp"
                /protein_id="AAS51373.1"
                /db_xref="GI:44981623"
                /translation="MAKHVEYSGRLRLHKFLSEPAELMAFKSAAPHFLHLQKKKYRKQ
                RRTLRRYFADDSFRYLPRDNAPKKSIAKYWSYRHDLFSLIDQGNIHLTEELWYSVTPE
                RVAKFTAAFIEACLPGATTILDVFCGGGGNVVHFAKVFQKVYGVELKLEHLYCTYRNA
                EAYGVADRVWLKHGDWLQLAARGRFERVHVDCVFASPPWGGIDYTKTQAYDLETQLQP
                AGLSQLLSSFLRISANVVLFLPKNSNLQQLSEITRSLLGPRAKCKVLYARENGRDKGL
                LCFWGEAFYNFDAASPANSGERLAARLQEAHDLYG"
gene            <9605. .>11434
                /locus_tag="ACR148W"
mRNA            <9605. .>11434
                /locus_tag="ACR148W"
                /product="ACR148Wp"
CDS             9605. .11434
                /locus_tag="ACR148W"
                /note="syntenic homolog of Saccharomyces cerevisiae
                YPL158C"
                /codon_start=1
                /evidence=not_experimental
                /product="ACR148Wp"
                /protein_id="AAS51374.1"
                /db_xref="GI:44981624"
                /translation="MPIGKLRRSKAISFRGPASPSLKFKSFETLQGKPLGEMGQNGHK
```

```
  Query Match             27.7%;  Score 18;  DB 8;  Length 306790;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         39 GAAACCCTGCTGATGGAC 56
              ||||||||||||||||||
Db       9493 GAAACCCTGCTGATGGAC 9510
```

```
RESULT 15
BX321857/c
LOCUS       BX321857               313050 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Nitrosomonas europaea ATCC 19718, complete genome; segment 2/10.
ACCESSION   BX321857 AL954747
VERSION     BX321857.1  GI:30138174
KEYWORDS    complete genome.
SOURCE      Nitrosomonas europaea ATCC 19718
  ORGANISM  Nitrosomonas europaea ATCC 19718
            Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
            Nitrosomonadaceae; Nitrosomonas.
REFERENCE   1
  AUTHORS   Chain,P., Lamerdin,J., Larimer,F., Regala,W., Land,M., Hauser,L.,
            Hooper,A., Klotz,M., Norton,J., Sayavedra-Soto,L., Arciero,D.,
            Hommes,N., Whittaker,M. and Arp,D.
  TITLE     Complete genome sequence of the ammonia-oxidizing bacterium and
            obligate chemolithoautotroph Nitrosomonas europaea
  JOURNAL   J. Bacteriol. 185 (9), 2759-2773 (2003)
  MEDLINE   22586410
   PUBMED   12700255
REFERENCE   2  (bases 1 to 313050)
  AUTHORS   Larimer,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas
            genome consortium, the DOE Joint Genome Institute, Production
            Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
            USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
            1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
            larimerfw@ornl.gov
  REMARK    Nitrosomonas genome consortium
FEATURES             Location/Qualifiers
     source          1. .313050
                     /organism="Nitrosomonas europaea ATCC 19718"
                     /mol_type="genomic DNA"
                     /strain="ATCC 19718"
                     /db_xref="taxon:228410"
     gene            complement(132. .503)
                     /locus_tag="NE0270"
     CDS             complement(132. .503)
                     /locus_tag="NE0270"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD84181.1"
                     /db_xref="GI:30138175"
                     /translation="MPEQITRKASPFCLRLTPEERTLLEREAAGLPLGEYIRQQVFDE
                     NRVKRRSRNKQPVKDHRLLSQLLGELGRSRLANNLNQLARAANCGLLNLTPEVKTSLL
                     NACADIRHIRETLMKSLGLNR"
     gene            complement(885. .1253)
                     /locus_tag="NE0271"
     CDS             complement(885. .1253)
```

```
                /locus_tag="NE0271"
                /function="InterPro IPR002559"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Transposase IS4 family"
                /protein_id="CAD84182.1"
                /db_xref="GI:30138176"
                /translation="MFSLSPGQAGDAPEGRKLLKTLENKGFSDTHVIMDKAYEGDETR
                QLVLDLGMIPVVPPKANRVSPWEYDVEMYKKRNEVERLFRRLKRFRRIFSRFDKLDSV
                FRFFIHFALIVDHLISVNRP"
misc_feature    complement(887. .1698)
                /note="ISNE5"
gene            complement(1253. .1660)
                /locus_tag="NE0272"
CDS             complement(1253. .1660)
                /locus_tag="NE0272"
                /function="COGs COG3293"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="possible transposase"
                /protein_id="CAD84183.1"
                /db_xref="GI:30138177"
                /translation="MEISASQFKLIENLLPIQRGNVTLSNLEVLNAILYVAEHGCKWR
                GLPVKFGNWHSIYTRANRWARNGVLDRVFLALQQNKLIQLEVDHMSLDSTIVKVHPDG
                TGALKKTVFKLSVNHEGAGLPKFTWSRQMPEPR"
gene            complement(1853. .2488)
                /gene="cyc"
                /locus_tag="NE0273"
CDS             complement(1853. .2488)
                /gene="cyc"
                /locus_tag="NE0273"
                /function="InterPro IPR000345:IPR002329:IPR003088 COGs
                COG2863"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Cytochrome c, class IC:Cytochrome c, class I"
                /protein_id="CAD84184.1"
                /db_xref="GI:30138178"
                /translation="MKFIGTTVAAIALTFSGFVFAESPAAGDVEKGKEIAAGICAGCH
                NADGNSAIPLYPILAGQYPGYIAKQLNDFKVVEGETVKRDNQIMAPMVATLSQDDIEN
                LAAYYSQQKPQPGTASDASLVETGKKLYQGGNLENSIPACSSCHSPNGQGIPPHYPRI
                DGQHPAYTLSQLQAFRQGTRKNDTNNTMQTIVSRMSEQEMQAVSEYIATLR"
misc_feature    complement(2423. .2488)
                /gene="cyc"
                /locus_tag="NE0273"
                /note="Signal predicted by SignalP 2.0 HMM (Signal peptide
                probabilty 1.000) with cleavage site probability 0.999 at
                residue 22"
gene            2595. .3197
                /locus_tag="NE0274"
CDS             2595. .3197
                /locus_tag="NE0274"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD84185.1"
                /db_xref="GI:30138179"
                /translation="MRKFSVFLLLANIFVVFYLHGRPDDNLPAQIALIHSEKIELLPA
                KVACLKWENLIGPVVQHVRVEISKWESGQDHITEISRGEVTVHWVHIPPLRNARETAK
                QIEQLKKSGISYLHIQENADSPWHNAISLAILPDDSDVAALVEELKGKGVERIMDSEQ
                VLEQFEFDIRNPTEQITESVRQLAQQFPETKLEVTECSRL"
misc_feature    2595. .2660
                /locus_tag="NE0274"
                /note="Signal predicted by SignalP 2.0 HMM (Signal peptide
                probabilty 0.892) with cleavage site probability 0.890 at
                residue 22"
gene            3311. .4129
                /gene="map"
                /locus_tag="NE0275"
CDS             3311. .4129
                /gene="map"
                /locus_tag="NE0275"
                /EC_number="3.4.11.18"
                /function="InterPro IPR000994:IPR001714:IPR002467 COGs
                COG0024"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Methionine aminopeptidase"
                /protein_id="CAD84186.1"
                /db_xref="GI:30138180"
                /translation="MQAGAGVVIHIKTAPEIETMRIAGRLASEVLDYIEPYVVPGVTT
                GELDELCHRYMVDVQKTIPAPLNYAPPGYSPYPKSICTSVNHQVCHGVPGDKKLKDGD
                VVNLDITVIYEGYHGDTSRMYYVGEPSIQAKRLCELTYEAMWRGIEEIRPGKHLGDIG
                HAIQRLAEGAGYSVVREFCGHGIGAKFHEDPQVLHYGRAGTGIELKPGMIFTVEPMIN
                AGKAAIKQLPDGWTVITKDHSLSAQWEHTILVTDESYEVLTVSSGSPARPAYRF"
gene            4197. .4922
                /gene="rph"
                /locus_tag="NE0276"
CDS             4197. .4922
                /gene="rph"
                /locus_tag="NE0276"
                /EC_number="2.7.7.56"
                /function="InterPro IPR001247:IPR002381 COGs COG0689"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="3' exoribonuclease family"
                /protein_id="CAD84187.1"
                /db_xref="GI:30138181"
                /translation="MPRCNNRAPAQMRPVRIIRHYVRHAEGSVLIEYGETRVICTASV
                IEKVPPFLKGAGQGWLTAEYGMLPRSTGERMQREAAKGKQSGRTMEIQRLIGRALRSI
                LDLEKLGERTIQMDCDVIQADGGTRTASITGAFVALYDAIDYLRAERMISQNPIRDHV
                AAVSVGILKGQPLLDLDYLEDSGCDTDLNVVMTGSLGLVEVQGTAEKVVFSRQELDVM
                LNMAQQGLQELFDVQRKALETVA"
gene            4960. .5571
                /locus_tag="NE0277"
CDS             4960. .5571
                /locus_tag="NE0277"
                /function="InterPro IPR002637 COGs COG0127"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Ham1 family"
                /protein_id="CAD84188.1"
                /db_xref="GI:30138182"
                /translation="MNKIVIASNNAGKLAEISRLLAPLGIEVVTQSSLGVTEADEPHM
                TFVENALAKARHASLATGLPALADDSGICVSALRGDPGVFSARYAGEPRSDERNNRKL
                VEALHGQSDRRAYYYCVIVLLRHGQDPQPVIIEDTWRGEIIAEPIGQGGFGYDPHFFL
                PELGKTAAELSIEEKNRISHRGKALARLVQMLSENETVPVVPV"
gene            5608. .6807
                /locus_tag="NE0278"
CDS             5608. .6807
                /locus_tag="NE0278"
                /EC_number="1.-.-.-"
                /function="InterPro IPR003401:IPR004559 COGs COG0635"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Coproporphyrinogen III oxidase and related FeS
                oxidoreductases"
                /protein_id="CAD84189.1"
                /db_xref="GI:30138183"
                /translation="MPVLSVPDSKSPHLAELPPLGLYIHLPWCVRKCPYCDFNSHAIN
                SGGDLPESEYVTALKRDLELALPLIWGRPVTSVFFGGGTPSLFSATAIDAILSHVRML
                LPLAPGVEITLEANPGTLEAQKFTDFHSAGINRLSIGIQSFNSRHLQALGRIHDGAEA
                LRAAELALQHFSNVNLDLMYALPGQTRDEALQDIETACNRGVAHISAYHLTIEPNTPF
                YRYPPQLPDEDTSADMQVMIEHILAERGYRHYETSAFAQPGKPCLHNMNYWQYGDYLG
                IGAGAHSKISFRDRIIRQMRHKHPRQYLEQAMAGNACLENLLATEQEVGPQDRIFEFM
                MNALRLTEGFDPGLFHTRTGLTIASIQKSLAEAEQRGLIEWQHDCIRPTPEGRRFLND
                LLEIFLS"
gene            6979. .8820
```

/gene="pcoA"
/locus_tag="NE0279"
CDS 6979. .8820
/gene="pcoA"

Query Match 27.7%; Score 18; DB 1; Length 313050;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         37 TCGAAACCCTGCTGATGG 54
              ||||||||||||||||||
Db      81964 TCGAAACCCTGCTGATGG 81947
```

Search completed: December 22, 2004, 23:36:25
Job time : 929.412 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:14:28 ; Search time 229.412 Seconds
(without alignments)
1487.336 Million cell updates/sec

Title: US-09-898-616A-1
Perfect score: 65
Sequence: 1 gatccatggaaatctgcccg.....tgctgatggacaccccgtcc 65

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	65	100.0	65	9	ABZ58370	Abz58370 Human ute
	2	65	100.0	65	12	ADL27626	Adl27626 Recombina
c	3	23	35.4	60	9	ABZ58376	Abz58376 Human ute
c	4	23	35.4	60	9	ABZ58377	Abz58377 Human ute
c	5	23	35.4	60	12	ADL27633	Adl27633 Recombina
c	6	23	35.4	60	12	ADL27632	Adl27632 Recombina
	7	17	26.2	36	6	ABK10937	Abk10937 Primer IR
	8	17	26.2	36	6	ABK47610	Abk47610 Interleuk
c	9	17	26.2	990	5	AAH66339	Aah66339 C glutami
c	10	17	26.2	1113	4	AAF71853	Aaf71853 Corynebac
c	11	17	26.2	1113	4	AAS96115	Aas96115 C. glutam
	12	17	26.2	2055	5	AAS74960	Aas74960 DNA encod
	13	17	26.2	2877	8	ACA43988	Aca43988 Prokaryot
	14	17	26.2	2892	4	AAS54085	Aas54085 Pseudomon
	15	17	26.2	2892	8	ACA42199	Aca42199 Prokaryot
	16	17	26.2	2922	8	ACA45765	Aca45765 Prokaryot
	17	17	26.2	2997	11	ABD04275	Abd04275 Pseudomon
c	18	17	26.2	3009	11	ABD04487	Abd04487 Pseudomon
c	19	17	26.2	7642	10	ADD44755	Add44755 Human gen
c	20	17	26.2	349980	5	AAH68528	Aah68528 C glutami
c	21	16	24.6	139	12	ACH90361	Ach90361 Human gen
c	22	16	24.6	230	8	ABZ19128	Abz19128 Group III
	23	16	24.6	277	8	ABZ18853	Abz18853 Group III
c	24	16	24.6	401	4	AAK95547	Aak95547 Human neu
c	25	16	24.6	401	4	AAK97040	Aak97040 Human neu
c	26	16	24.6	401	6	ABT00317	Abt00317 Human neu
c	27	16	24.6	401	6	ABT01810	Abt01810 Human neu
	28	16	24.6	426	8	ACA43852	Aca43852 Prokaryot
c	29	16	24.6	465	8	ABZ20021	Abz20021 Group III
c	30	16	24.6	490	8	ABZ19919	Abz19919 Group III
c	31	16	24.6	513	12	ACH76661	Ach76661 Human gen
	32	16	24.6	518	8	ABZ19817	Abz19817 Group III
c	33	16	24.6	543	6	ABK73249	Abk73249 Bacillus
	34	16	24.6	629	8	ABZ19560	Abz19560 Group III
	35	16	24.6	646	8	ABZ19299	Abz19299 Group III
	36	16	24.6	894	11	ACH95360	Ach95360 Klebsiell
c	37	16	24.6	1044	8	ABZ73819	Abz73819 Secreted
c	38	16	24.6	1044	10	ADC20635	Adc20635 Human sec
c	39	16	24.6	1044	10	ABT16890	Abt16890 Human sec
c	40	16	24.6	1044	10	ABZ67403	Abz67403 Human sec
c	41	16	24.6	1052	3	AAC59250	Aac59250 Human sec
c	42	16	24.6	1052	8	ABZ73338	Abz73338 Secreted
c	43	16	24.6	1052	10	ADC20099	Adc20099 Human sec
c	44	16	24.6	1052	10	ABT16781	Abt16781 Human sec
c	45	16	24.6	1052	10	ABZ66948	Abz66948 Human sec

ALIGNMENTS

RESULT 1
ABZ58370
ID ABZ58370 standard; DNA; 65 BP.
XX
AC ABZ58370;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 1.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Claim 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 1, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and

CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 65 BP; 13 A; 22 C; 14 G; 16 T; 0 U; 0 Other;

Query Match 100.0%; Score 65; DB 9; Length 65;
Best Local Similarity 100.0%; Pred. No. 2e-25;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy 61 CGTCC 65
|||||
Db 61 CGTCC 65

RESULT 2
ADL27626
ID ADL27626 standard; DNA; 65 BP.
XX
AC ADL27626;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, coding oligonucleotide #1.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Claim 1; SEQ ID NO 1; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a coding strand oligonucleotide used to
CC construct the synthetic rhUG gene.
XX
SQ Sequence 65 BP; 13 A; 22 C; 14 G; 16 T; 0 U; 0 Other;

Query Match 100.0%; Score 65; DB 12; Length 65;
Best Local Similarity 100.0%; Pred. No. 2e-25;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy 61 CGTCC 65
|||||
Db 61 CGTCC 65

RESULT 3
ABZ58376/c
ID ABZ58376 standard; DNA; 60 BP.
XX
AC ABZ58376;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 7.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic

PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 7, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 35.4%; Score 23; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 CCCTGCTGATGGACACCCCGTCC 65
||||||||||||||||||||||||
Db 60 CCCTGCTGATGGACACCCCGTCC 38

RESULT 4
ABZ58377/c
ID ABZ58377 standard; DNA; 60 BP.
XX
AC ABZ58377;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 8.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 8, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 35.4%; Score 23; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 CCCTGCTGATGGACACCCCGTCC 65
||||||||||||||||||||||||
Db 60 CCCTGCTGATGGACACCCCGTCC 38

RESULT 5
ADL27633/c
ID ADL27633 standard; DNA; 60 BP.
XX
AC ADL27633;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #4.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 8; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included

CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 35.4%; Score 23; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 CCCTGCTGATGGACACCCCGTCC 65
||||||||||||||||||||||||
Db 60 CCCTGCTGATGGACACCCCGTCC 38

RESULT 6
ADL27632/c
ID ADL27632 standard; DNA; 60 BP.
XX
AC ADL27632;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #3.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 7; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 35.4%; Score 23; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 CCCTGCTGATGGACACCCCGTCC 65
||||||||||||||||||||||||
Db 60 CCCTGCTGATGGACACCCCGTCC 38

RESULT 7
ABK10937
ID ABK10937 standard; DNA; 36 BP.
XX
AC ABK10937;
XX
DT 20-MAY-2002 (first entry)
XX
DE Primer IRAP5 relating to modified retroviral vector invention.
XX
KW Human; retrovirus vector; elongation factor 1-alpha; gene therapy;
KW EF 1alpha; Murine Leukaemia Virus; MLV; multi-cloning site; MCS; primer;
KW IRAP5; ss.
XX
OS Unidentified.
XX
PN KR2001069245-A.
XX
PD 25-JUL-2001.
XX
PF 08-SEP-2000; 2000KR-00053613.
XX
PR 08-SEP-2000; 2000KR-00053613.
XX
PA (VIRO-) VIROMED LTD.
XX
PI Kim SY, Lee JT, Yoo SS;

XX
DR WPI; 2002-065240/09.
XX
PT Retroviral vectors useful in gene therapy, containing no viral coding
PT sequences but which includes a human elongation factor 1 alpha non-coding
PT fragment.
XX
PS Disclosure; Page 31; 32pp; Korean.
XX
CC The present invention relates to a retrovirus vector containing modified
CC non-coding sequences derived from the human elongation factor 1-alpha
CC gene (EF 1alpha, bases +773 to +1006), which can be effectively used in
CC gene therapy due to it's high stability and expression. The Murine
CC Leukaemia Virus (MLV)-derived retrovirus vector is modified by complete
CC removal of the gag, env and pol genes to improve gene expression and
CC virus productability. It is also modified to contain the non-coding
CC portion of the EF 1alpha gene upstream of a multi-cloning site (MCS). The
CC present sequence represents a primer used in the methods of the present
CC invention
XX
SQ Sequence 36 BP; 9 A; 11 C; 10 G; 6 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCCATGGAAATCTGC 17
|||||||||||||||||
Db 2 GATCCATGGAAATCTGC 18

RESULT 8
ABK47610
ID ABK47610 standard; DNA; 36 BP.
XX
AC ABK47610;
XX
DT 24-SEP-2002 (first entry)
XX
DE Interleukin-1 receptor antagonist (IL-1ra), PCR primer IRAP5'.
XX
KW Human; elongation factor 1-alpha; retroviral vector; EF1a;
KW murine leukaemia virus vector; MLV; elongation factor 1 alpha;
KW gene therapy; interleukin-1 receptor antagonist; IL-1ra; PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200220810-A1.
XX
PD 14-MAR-2002.
XX
PF 08-SEP-2001; 2001WO-KR001515.
XX
PR 08-SEP-2000; 2000KR-00053613.
XX
PA (VIRO-) VIROMED LTD.
XX
PI Kim SY, Yu SS, Lee JT;
XX
DR WPI; 2002-065240/41.
XX
PT Retroviral vectors useful in gene therapy, containing no viral coding
PT sequences but which includes a human elongation factor 1 alpha non-coding
PT fragment.
XX
PS Example 5; Page 45; 47pp; English.
XX
CC The invention relates to a novel retroviral vector, derived from a murine
CC leukaemia virus (MLV) vector, which lacks viral coding sequences but
CC which includes: (i) part of the non-coding sequence (I) of elongation
CC factor 1 alpha (EF1a), as heterologous, gene-derived sequence, inserted
CC upstream of the multiple cloning site; and (ii) a mutation, downstream of
CC the splice acceptor, within (I). The retroviral vectors are useful as
CC gene therapy vectors. They are safe, i.e. they can not form replication-
CC competent retroviruses by homologous recombination; the heterologous
CC intron ensures efficient expression of foreign genes and the specified
CC mutation means that a proper balance between splicing efficiency and
CC viral titre is maintained. Also a heterologous promoter can be inserted
CC to increase viral titre in packaging lines derived from human cells. An
CC additional promoter (or internal ribosome entry site) can be provided to
CC allow expression of two or more genes. The present sequence represents a
CC PCR primer used to make the retroviral vector of the invention
XX
SQ Sequence 36 BP; 9 A; 11 C; 10 G; 6 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCCATGGAAATCTGC 17
|||||||||||||||||
Db 2 GATCCATGGAAATCTGC 18

RESULT 9
AAH66339/c
ID AAH66339 standard; DNA; 990 BP.
XX
AC AAH66339;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 1374.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG91120.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 8; SEQ ID NO 1374; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX

```
SQ   Sequence 990 BP; 196 A; 239 C; 310 G; 245 T; 0 U; 0 Other;

  Query Match             26.2%;  Score 17;  DB 5;  Length 990;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGC 17
              |||||||||||||||||
Db        688 GATCCATGGAAATCTGC 672


RESULT 10
AAF71853/c
ID   AAF71853 standard; DNA; 1113 BP.
XX
AC   AAF71853;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:201.
XX
KW   Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW   fine chemical production; microorganism; organic acid; nucleoside;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW   lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW   carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100843-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB000923.
XX
PR   25-JUN-1999;   99US-0141031P.
PR   01-JUL-1999;   99DE-01030476.
PR   02-JUL-1999;   99US-0142101P.
PR   08-JUL-1999;   99DE-01031415.
PR   08-JUL-1999;   99DE-01031418.
PR   08-JUL-1999;   99DE-01031419.
PR   08-JUL-1999;   99DE-01031420.
PR   08-JUL-1999;   99DE-01031424.
PR   08-JUL-1999;   99DE-01031428.
PR   08-JUL-1999;   99DE-01031434.
PR   08-JUL-1999;   99DE-01031435.
PR   08-JUL-1999;   99DE-01031443.
PR   08-JUL-1999;   99DE-01031453.
PR   08-JUL-1999;   99DE-01031457.
PR   08-JUL-1999;   99DE-01031465.
PR   08-JUL-1999;   99DE-01031478.
PR   08-JUL-1999;   99DE-01031510.
PR   08-JUL-1999;   99DE-01031541.
PR   08-JUL-1999;   99DE-01031573.
PR   08-JUL-1999;   99DE-01031592.
PR   08-JUL-1999;   99DE-01031632.
PR   08-JUL-1999;   99DE-01031634.
PR   08-JUL-1999;   99DE-01031636.
PR   09-JUL-1999;   99DE-01032125.
PR   09-JUL-1999;   99DE-01032126.
PR   09-JUL-1999;   99DE-01032130.
PR   09-JUL-1999;   99DE-01032186.
PR   09-JUL-1999;   99DE-01032206.
PR   09-JUL-1999;   99DE-01032227.
PR   09-JUL-1999;   99DE-01032228.
PR   09-JUL-1999;   99DE-01032229.
PR   09-JUL-1999;   99DE-01032230.
PR   14-JUL-1999;   99DE-01032922.
PR   14-JUL-1999;   99DE-01032926.
PR   14-JUL-1999;   99DE-01032928.
PR   14-JUL-1999;   99DE-01033004.
PR   14-JUL-1999;   99DE-01033005.
PR   14-JUL-1999;   99DE-01033006.
PR   12-AUG-1999;   99US-0148613P.
PR   27-AUG-1999;   99DE-01040764.
PR   27-AUG-1999;   99DE-01040765.
PR   27-AUG-1999;   99DE-01040766.
PR   27-AUG-1999;   99DE-01040832.
PR   31-AUG-1999;   99DE-01041378.
PR   31-AUG-1999;   99DE-01041379.
PR   31-AUG-1999;   99DE-01041380.
PR   31-AUG-1999;   99DE-01041394.
PR   31-AUG-1999;   99DE-01041396.
PR   03-SEP-1999;   99DE-01042076.
PR   03-SEP-1999;   99DE-01042077.
PR   03-SEP-1999;   99DE-01042079.
PR   03-SEP-1999;   99DE-01042086.
PR   03-SEP-1999;   99DE-01042087.
PR   03-SEP-1999;   99DE-01042088.
PR   03-SEP-1999;   99DE-01042095.
PR   03-SEP-1999;   99DE-01042124.
PR   03-SEP-1999;   99DE-01042129.
PR   09-MAR-2000; 2000US-0187970P.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-137957/14.
DR   P-PSDB; AAB79734.
XX
PT   Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway
PT   proteins, useful for producing fine chemicals in microorganisms,
PT   including organic acids, nonproteinogenic amino acids, and purine and
PT   pyrimidine bases.
XX
PS   Claim 3; Page 446-447; 1737pp; English.
XX
CC   AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum MP
CC   nucleic acids are useful for the production of fine chemicals in
CC   microorganisms, including organic acids, nonproteinogenic amino acids,
CC   purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated
CC   and unsaturated fatty acids, diols, carbohydrates, aromatic compounds,
CC   vitamins, cofactors, polyketides and enzymes
XX
SQ   Sequence 1113 BP; 220 A; 271 C; 348 G; 274 T; 0 U; 0 Other;

  Query Match             26.2%;  Score 17;  DB 4;  Length 1113;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGC 17
              |||||||||||||||||
Db        788 GATCCATGGAAATCTGC 772


RESULT 11
AAS96115/c
ID   AAS96115 standard; DNA; 1113 BP.
XX
AC   AAS96115;
XX
DT   26-FEB-2002  (first entry)
XX
DE   C. glutamicum gene #40 encoding metabolic pathway protein.
XX
KW   Metabolic pathway protein; MP; lysine biosynthesis pathway;
KW   methionine biosynthesis pathway; large-scale production of fine chemical;
KW   Corynebacterium diphtheriae; diphtheria; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200166573-A2.
XX
PD   13-SEP-2001.
```

XX
PF 22-DEC-2000; 2000WO-IB002035.
XX
PR 09-MAR-2000; 2000US-0187970P.
PR 23-JUN-2000; 2000US-00606740.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI Kim J, Lee H, Hwang B;
XX
DR WPI; 2001-582269/65.
DR P-PSDB; AAU71905.
XX
PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium
PT glutamicum, useful for producing methionine and lysine in Corynebacterim
PT and Brevibacterium.
XX
PS Disclosure; Page 256-257; 316pp; English.
XX
CC The present invention relates to the isolation of novel Corynebacterium
CC glutamicum genes encoding metabolic pathway (MP) proteins (AAU71863-
CC AAU71922). The metabolic pathway proteins of the invention include
CC enzymes involved in the lysine and methionine biosynthetic pathways. The
CC polynucleotide sequences of the invention can be used for the large-scale
CC production and/or modulation of expression of fine chemicals such as
CC lysine and methionine. The sequences of the invention may be used to
CC identify C. glutamicum and related organisms e.g. C. diphtheriae in a
CC subject to detect diphtheria. AAS96073-AAS96132 represent C. glutamicum
CC genes encoding the novel metabolic pathway proteins of the invention
XX
SQ Sequence 1113 BP; 220 A; 271 C; 348 G; 274 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 4; Length 1113;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCCATGGAAATCTGC 17
||||||||||||||||
Db 788 GATCCATGGAAATCTGC 772

RESULT 12
AAS74960
ID AAS74960 standard; cDNA; 2055 BP.
XX
AC AAS74960;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10764.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG10773.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 10764; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2055 BP; 469 A; 581 C; 604 G; 401 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 5; Length 2055;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 AAACCCTGCTGATGGAC 56
||||||||||||||||
Db 98 AAACCCTGCTGATGGAC 114

RESULT 13
ACA43988
ID ACA43988 standard; DNA; 2877 BP.
XX
AC ACA43988;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #25645.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas putida.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.

DR P-PSDB; ABU40118.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 31858; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2877 BP; 629 A; 981 C; 802 G; 465 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 8; Length 2877;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 CCTGCTGATGGACACCC 60
|||||||||||||||||
Db 675 CCTGCTGATGGACACCC 691

RESULT 14
AAS54085
ID AAS54085 standard; DNA; 2892 BP.
XX
AC AAS54085;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #216.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU36226.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 7722; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2892 BP; 594 A; 1063 C; 824 G; 411 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 4; Length 2892;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 CCTGCTGATGGACACCC 60
|||||||||||||||||
Db 675 CCTGCTGATGGACACCC 691

RESULT 15
ACA42199
ID ACA42199 standard; DNA; 2892 BP.
XX
AC ACA42199;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #23856.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU38329.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 30069; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2892 BP; 594 A; 1063 C; 824 G; 411 T; 0 U; 0 Other;

Query Match 26.2%; Score 17; DB 8; Length 2892;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 44 CCTGCTGATGGACACCC 60
||||||||||||||||
Db 675 CCTGCTGATGGACACCC 691

Search completed: December 22, 2004, 22:44:07
Job time : 234.662 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 22:08:48 ; Search time 52.3529 Seconds
(without alignments)
882.496 Million cell updates/sec

Title: US-09-898-616A-1
Perfect score: 65
Sequence: 1 gatccatggaaatctgcccg.....tgctgatggacaccccgtcc 65

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	65	100.0	65	3	US-08-864-357F-6	Sequence 6, Appli
c	2	42	64.6	42	3	US-08-864-357F-13	Sequence 13, Appl
c	3	23	35.4	60	3	US-08-864-357F-12	Sequence 12, Appl
	4	17	26.2	2997	4	US-09-252-991A-2879	Sequence 2879, Ap
c	5	17	26.2	3009	4	US-09-252-991A-3091	Sequence 3091, Ap
	6	16	24.6	894	4	US-09-489-039A-1155	Sequence 1155, Ap
	7	16	24.6	1119	4	US-09-489-039A-3704	Sequence 3704, Ap
	8	16	24.6	1155	3	US-08-858-207A-101	Sequence 101, App
c	9	16	24.6	1434	4	US-09-023-655-481	Sequence 481, App
c	10	16	24.6	3911	4	US-09-566-921-61	Sequence 61, Appl
	11	16	24.6	4169	4	US-09-166-350-32	Sequence 32, Appl
c	12	15	23.1	301	4	US-09-513-999C-11687	Sequence 11687, A
	13	15	23.1	434	4	US-09-702-705-1590	Sequence 1590, Ap
	14	15	23.1	434	4	US-09-736-457-1590	Sequence 1590, Ap
	15	15	23.1	434	4	US-09-614-124B-1590	Sequence 1590, Ap
	16	15	23.1	434	4	US-09-671-325-1590	Sequence 1590, Ap
	17	15	23.1	434	4	US-09-658-824-1590	Sequence 1590, Ap
	18	15	23.1	472	4	US-09-702-705-273	Sequence 273, App
	19	15	23.1	472	4	US-09-702-705-342	Sequence 342, App
	20	15	23.1	472	4	US-09-736-457-273	Sequence 273, App
	21	15	23.1	472	4	US-09-736-457-342	Sequence 342, App
	22	15	23.1	472	4	US-09-614-124B-273	Sequence 273, App
	23	15	23.1	472	4	US-09-614-124B-342	Sequence 342, App
	24	15	23.1	472	4	US-09-671-325-273	Sequence 273, App
	25	15	23.1	472	4	US-09-671-325-342	Sequence 342, App
	26	15	23.1	472	4	US-09-589-184-273	Sequence 273, App
	27	15	23.1	472	4	US-09-589-184-342	Sequence 342, App
	28	15	23.1	472	4	US-09-658-824-273	Sequence 273, App
	29	15	23.1	472	4	US-09-658-824-342	Sequence 342, App
	30	15	23.1	1062	4	US-09-248-796A-4823	Sequence 4823, Ap
	31	15	23.1	1303	4	US-09-566-921-117	Sequence 117, App
	32	15	23.1	1578	4	US-09-489-039A-2526	Sequence 2526, Ap
c	33	15	23.1	2001	4	US-09-375-318-35	Sequence 35, Appl
	34	15	23.1	3032	4	US-09-799-451-152	Sequence 152, App
	35	15	23.1	3463	3	US-09-189-462-3	Sequence 3, Appli
	36	15	23.1	3463	4	US-09-863-040-3	Sequence 3, Appli
	37	15	23.1	4451	3	US-08-717-294-42	Sequence 42, Appl
	38	15	23.1	5323	4	US-09-799-451-726	Sequence 726, App
	39	15	23.1	8355	3	US-08-406-030A-23	Sequence 23, Appl
c	40	15	23.1	9381	3	US-09-453-702B-7	Sequence 7, Appli
c	41	15	23.1	12848	3	US-09-453-702B-252	Sequence 252, App
	42	15	23.1	38155	3	US-09-453-702B-79	Sequence 79, Appl
c	43	15	23.1	42325	4	US-08-311-731A-131	Sequence 131, App
c	44	15	23.1	48908	3	US-09-453-702B-137	Sequence 137, App
	45	15	23.1	1664976	4	US-08-916-421B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-864-357F-6
; Sequence 6, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 65
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-6

Query Match 100.0%; Score 65; DB 3; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65
```

RESULT 2
US-08-864-357F-13/c
; Sequence 13, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 42
; TYPE: DNA

```
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-08-864-357F-13

  Query Match             64.6%;  Score 42;  DB 3;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-14;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 42
              |||||||||||||||||||||||||||||||||||||||||
Db         42 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 1


RESULT 3
US-08-864-357F-12/c
; Sequence 12, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
;  APPLICANT:  Claragen, Inc. & NIH
;  TITLE OF INVENTION:   Use of Recombinant Human Uteroglobin in Treatment of Inflammatc
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE:   116142/2
;  CURRENT APPLICATION NUMBER: US/08/864,357F
;  CURRENT FILING DATE:  1997-05-28
;  NUMBER OF SEQ ID NOS:   22
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-08-864-357F-12

  Query Match             35.4%;  Score 23;  DB 3;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 4
US-09-252-991A-2879
; Sequence 2879, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2879
;   LENGTH: 2997
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2879

  Query Match             26.2%;  Score 17;  DB 4;  Length 2997;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          44 CCTGCTGATGGACACCC 60
               |||||||||||||||||
Db         780 CCTGCTGATGGACACCC 796


RESULT 5
US-09-252-991A-3091/c
; Sequence 3091, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 3091
;   LENGTH: 3009
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3091

  Query Match             26.2%;  Score 17;  DB 4;  Length 3009;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          44 CCTGCTGATGGACACCC 60
               |||||||||||||||||
Db        2428 CCTGCTGATGGACACCC 2412


RESULT 6
US-09-489-039A-1155
; Sequence 1155, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 1155
;   LENGTH: 894
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1155

  Query Match             24.6%;  Score 16;  DB 4;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          40 AAACCCTGCTGATGGA 55
               ||||||||||||||||
Db         527 AAACCCTGCTGATGGA 542


RESULT 7
US-09-489-039A-3704
; Sequence 3704, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
```

```
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 3704
;   LENGTH: 1119
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3704

  Query Match             24.6%;  Score 16;  DB 4;  Length 1119;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 CTGCTGATGGACACCC 60
               ||||||||||||||||
Db         937 CTGCTGATGGACACCC 952


RESULT 8
US-08-858-207A-101
; Sequence 101, Application US/08858207A
; Patent No. 6348328
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Stodola, Robert
;    TITLE OF INVENTION:  No. 6348328el Compounds
;    NUMBER OF SEQUENCES:  552
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,207A
;      FILING DATE:  09-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017670
;      FILING DATE:  14-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  101:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1155 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-858-207A-101

  Query Match             24.6%;  Score 16;  DB 3;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           3 TCCATGGAAATCTGCC 18
               ||||||||||||||||
Db        1011 TCCATGGAAATCTGCC 1026


RESULT 9
US-09-023-655-481/c
; Sequence 481, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  481:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1434 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  CONNTUT01
;      CLONE:  1908804
US-09-023-655-481

  Query Match             24.6%;  Score 16;  DB 4;  Length 1434;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          14 CTGCCCGTCTTTCCAG 29
               ||||||||||||||||
Db          26 CTGCCCGTCTTTCCAG 11


RESULT 10
US-09-566-921-61/c
; Sequence 61, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
;  APPLICANT:  Loring, Jeanne F.
;  APPLICANT:  Tingley, Debora W.
;  APPLICANT:  Edwards, Carla M.
;  TITLE OF INVENTION:  GENES EXPRESSED IN ALZHEIMER'S DISEASE
;  FILE REFERENCE:  PA-0024 US
;  CURRENT APPLICATION NUMBER: US/09/566,921
;  CURRENT FILING DATE:  2000-05-05
;  NUMBER OF SEQ ID NOS:  138
;  SOFTWARE:  PERL Program
```

; SEQ ID NO 61
; LENGTH: 3911
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6682888 166400.9
US-09-566-921-61

Query Match 24.6%; Score 16; DB 4; Length 3911;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 CTGCCCGTCTTTCCAG 29
||||||||||||||||
Db 2494 CTGCCCGTCTTTCCAG 2479

RESULT 11
US-09-166-350-32
; Sequence 32, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 4169
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-32

Query Match 24.6%; Score 16; DB 4; Length 4169;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 CCCTGCTGATGGACAC 58
||||||||||||||||
Db 3093 CCCTGCTGATGGACAC 3108

RESULT 12
US-09-513-999C-11687/c
; Sequence 11687, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 11687
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 244
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 274
; OTHER INFORMATION: r=a or g
US-09-513-999C-11687

Query Match 23.1%; Score 15; DB 4; Length 301;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 CCCGTCTTTCCAGCG 31
|||||||||||||||
Db 32 CCCGTCTTTCCAGCG 18

RESULT 13
US-09-702-705-1590
; Sequence 1590, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1590
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(434)
; OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1590

Query Match 23.1%; Score 15; DB 4; Length 434;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 GAAACCCTGCTGATG 53
|||||||||||||||
Db 20 GAAACCCTGCTGATG 34

RESULT 14
US-09-736-457-1590
; Sequence 1590, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane

```
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1590
;   LENGTH: 434
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(434)
;   OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1590

  Query Match             23.1%;  Score 15;  DB 4;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         39 GAAACCCTGCTGATG 53
              |||||||||||||||
Db         20 GAAACCCTGCTGATG 34


RESULT 15
US-09-614-124B-1590
; Sequence 1590, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C9
;  CURRENT APPLICATION NUMBER: US/09/614,124B
;  CURRENT FILING DATE:  2001-07-11
;  NUMBER OF SEQ ID NOS: 1668
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1590
;   LENGTH: 434
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(434)
;   OTHER INFORMATION: n = A,T,C or G
US-09-614-124B-1590

  Query Match             23.1%;  Score 15;  DB 4;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         39 GAAACCCTGCTGATG 53
              |||||||||||||||
Db         20 GAAACCCTGCTGATG 34


Search completed: December 23, 2004, 01:33:34
Job time : 55.3529 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 23:36:53 ; Search time 911.177 Seconds
(without alignments)
397.214 Million cell updates/sec

Title: US-09-898-616A-1
Perfect score: 65
Sequence: 1 gatccatggaaatctgcccg.....tgctgatggacaccccgtcc 65

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4105333 seqs, 2784095677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8210666

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	100.0	65	9	US-09-861-688-6	Sequence 6, Appli
2	65	100.0	65	10	US-09-898-616A-1	Sequence 1, Appli
3	65	100.0	65	15	US-10-187-498A-1	Sequence 1, Appli
4	65	100.0	65	16	US-10-647-371-5	Sequence 5, Appli
c 5	42	64.6	42	9	US-09-861-688-13	Sequence 13, Appl
c 6	42	64.6	42	16	US-10-647-371-12	Sequence 12, Appl
c 7	23	35.4	60	9	US-09-861-688-12	Sequence 12, Appl
c 8	23	35.4	60	10	US-09-898-616A-7	Sequence 7, Appli
c 9	23	35.4	60	10	US-09-898-616A-8	Sequence 8, Appli
c 10	23	35.4	60	15	US-10-187-498A-7	Sequence 7, Appli
c 11	23	35.4	60	15	US-10-187-498A-8	Sequence 8, Appli
c 12	23	35.4	60	16	US-10-647-371-11	Sequence 11, Appl
13	17	26.2	36	15	US-10-129-422-10	Sequence 10, Appl
14	17	26.2	534	18	US-10-425-115-112555	Sequence 112555,
15	17	26.2	647	16	US-10-424-599-87144	Sequence 87144, A
c 16	17	26.2	793	16	US-10-424-599-108961	Sequence 108961,
17	17	26.2	877	16	US-10-424-599-87143	Sequence 87143, A
c 18	17	26.2	990	9	US-09-738-626-1374	Sequence 1374, Ap
c 19	17	26.2	1113	10	US-09-746-660A-85	Sequence 85, Appl
20	17	26.2	1404	15	US-10-369-493-37571	Sequence 37571, A
c 21	17	26.2	1685	18	US-10-425-115-112556	Sequence 112556,
22	17	26.2	1887	16	US-10-425-114-30051	Sequence 30051, A
23	17	26.2	1908	16	US-10-425-114-12685	Sequence 12685, A
24	17	26.2	2645	16	US-10-424-599-87145	Sequence 87145, A
25	17	26.2	2877	16	US-10-282-122A-31858	Sequence 31858, A
26	17	26.2	2892	9	US-09-815-242-7722	Sequence 7722, Ap
27	17	26.2	2892	15	US-10-369-493-37445	Sequence 37445, A
28	17	26.2	2892	16	US-10-282-122A-30069	Sequence 30069, A
29	17	26.2	2922	16	US-10-282-122A-33635	Sequence 33635, A
c 30	17	26.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
c 31	16	24.6	139	15	US-10-029-386-23556	Sequence 23556, A
c 32	16	24.6	401	9	US-09-795-668-346	Sequence 346, App
c 33	16	24.6	401	9	US-09-795-686-346	Sequence 346, App
c 34	16	24.6	401	9	US-09-946-807-346	Sequence 346, App
35	16	24.6	426	16	US-10-282-122A-31722	Sequence 31722, A
c 36	16	24.6	513	15	US-10-029-386-9856	Sequence 9856, Ap
37	16	24.6	526	13	US-10-027-632-141374	Sequence 141374,
38	16	24.6	526	13	US-10-027-632-141375	Sequence 141375,
39	16	24.6	526	15	US-10-027-632-141374	Sequence 141374,
40	16	24.6	526	15	US-10-027-632-141375	Sequence 141375,
c 41	16	24.6	543	9	US-09-974-300-540	Sequence 540, App
c 42	16	24.6	1434	16	US-10-641-643-481	Sequence 481, App
43	16	24.6	1447	16	US-10-264-237-304	Sequence 304, App
44	16	24.6	1599	15	US-10-369-493-34988	Sequence 34988, A
45	16	24.6	1732	10	US-09-764-891-1049	Sequence 1049, Ap

ALIGNMENTS

RESULT 1
US-09-861-688-6
; Sequence 6, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-6

Query Match 100.0%; Score 65; DB 9; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.3e-27;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65
```

```
RESULT 2
US-09-898-616A-1
; Sequence 1, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-1

  Query Match             100.0%;  Score 65;  DB 10;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-27;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65


RESULT 3
US-10-187-498A-1
; Sequence 1, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-1

  Query Match             100.0%;  Score 65;  DB 15;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-27;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65


RESULT 4
US-10-647-371-5
; Sequence 5, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-5

  Query Match             100.0%;  Score 65;  DB 16;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-27;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAACCCTGCTGATGGACACCC 60

Qy         61 CGTCC 65
              |||||
Db         61 CGTCC 65


RESULT 5
US-09-861-688-13/c
; Sequence 13, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
;  TITLE OF INVENTION:  Inflammatory and
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE: 116142/2
;  CURRENT APPLICATION NUMBER: US/09/861,688
;  CURRENT FILING DATE:  2001-05-21
;  PRIOR APPLICATION NUMBER: 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 42
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer sequence
```

US-09-861-688-13

Query Match 64.6%; Score 42; DB 9; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 42
              ||||||||||||||||||||||||||||||||||||||||||
Db         42 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 1
```

RESULT 6
US-10-647-371-12/c
; Sequence 12, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
; TITLE OF INVENTION: and Fibrotic Conditions
; FILE REFERENCE: 116142-85
; CURRENT APPLICATION NUMBER: US/10/647,371
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: 09/549,926
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 42
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer sequence
US-10-647-371-12

Query Match 64.6%; Score 42; DB 16; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 42
              ||||||||||||||||||||||||||||||||||||||||||
Db         42 GATCCATGGAAATCTGCCCGTCTTTCCAGCGTGTTATCGAAA 1
```

RESULT 7
US-09-861-688-12/c
; Sequence 12, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-12

Query Match 35.4%; Score 23; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         43 CCCTGCTGATGGACACCCCGTCC 65
              |||||||||||||||||||||||
Db         60 CCCTGCTGATGGACACCCCGTCC 38
```

RESULT 8
US-09-898-616A-7/c
; Sequence 7, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
; APPLICANT: Claragen Inc.
; APPLICANT: Pilon, Aprile L
; APPLICANT: Welch, Richard W
; TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
; TITLE OF INVENTION: Imflammatory and Fibrotic Conditions
; FILE REFERENCE: 116142/00170
; CURRENT APPLICATION NUMBER: US/09/898,616A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Original amino acid sequence obtained from cDNA. Current sumitte
; OTHER INFORMATION: d sequence maximized for expression in E. coli.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Original amino acid sequence obtained from cDNA. Current sumitte
; OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-7

Query Match 35.4%; Score 23; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         43 CCCTGCTGATGGACACCCCGTCC 65
              |||||||||||||||||||||||
Db         60 CCCTGCTGATGGACACCCCGTCC 38
```

RESULT 9
US-09-898-616A-8/c
; Sequence 8, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
; APPLICANT: Claragen Inc.
; APPLICANT: Pilon, Aprile L
; APPLICANT: Welch, Richard W
; TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
; TITLE OF INVENTION: Imflammatory and Fibrotic Conditions
; FILE REFERENCE: 116142/00170
; CURRENT APPLICATION NUMBER: US/09/898,616A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Original amino acid sequence obtained from cDNA. Current sumitte
; OTHER INFORMATION: d sequence maximized for expression in E. coli.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Original amino acid sequence obtained from cDNA. Current sumitte
; OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-8

Query Match 35.4%; Score 23; DB 10; Length 60;

```
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 10
US-10-187-498A-7/c
; Sequence 7, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-7

  Query Match             35.4%;  Score 23;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 11
US-10-187-498A-8/c
; Sequence 8, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-8

  Query Match             35.4%;  Score 23;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 12
US-10-647-371-11/c
; Sequence 11, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-11

  Query Match             35.4%;  Score 23;  DB 16;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 0.0034;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          43 CCCTGCTGATGGACACCCCGTCC 65
               |||||||||||||||||||||||
Db          60 CCCTGCTGATGGACACCCCGTCC 38


RESULT 13
US-10-129-422-10
; Sequence 10, Application US/10129422
; Publication No. US20030166251A1
; GENERAL INFORMATION:
;  APPLICANT: KIM, Sun Young et al.; ViroMed Ltd.
;  TITLE OF INVENTION: High efficiency retroviral vector which contains genetically engir
;  TITLE OF INVENTION:  cellular
;  TITLE OF INVENTION:  non-coding sequence harboring splice acceptor
;  FILE REFERENCE: PCA10935/VML/PCT
;  CURRENT APPLICATION NUMBER: US/10/129,422
;  CURRENT FILING DATE:  2002-10-16
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: KOPATIN 1.5
; SEQ ID NO 10
;   LENGTH: 36
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer IRAP5'
US-10-129-422-10

  Query Match             26.2%;  Score 17;  DB 15;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 GATCCATGGAAATCTGC 17
               |||||||||||||||||
```

```
Db          2 GATCCATGGAAATCTGC 18


RESULT 14
US-10-425-115-112555
; Sequence 112555, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 112555
;   LENGTH: 534
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_34141C.1
US-10-425-115-112555

  Query Match             26.2%;  Score 17;  DB 18;  Length 534;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         12 ATCTGCCCGTCTTTCCA 28
              |||||||||||||||||
Db        454 ATCTGCCCGTCTTTCCA 470


RESULT 15
US-10-424-599-87144
; Sequence 87144, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 87144
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_49701C.1
US-10-424-599-87144

  Query Match             26.2%;  Score 17;  DB 16;  Length 647;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         33 GTTATCGAAACCCTGCT 49
              |||||||||||||||||
Db        413 GTTATCGAAACCCTGCT 429


Search completed: December 23, 2004, 05:19:24
Job time : 920.177 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:17:58 ; Search time 848.688 Seconds
(without alignments)
3343.258 Million cell updates/sec

Title: US-09-898-616A-2
Perfect score: 60
Sequence: 1 agctacgaagcagctatgga.....acatgcgtgaagcaggtgct 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60	100.0	60	6	AR160916.	AR160916 Sequence
c	2	37	61.7	60	6	AR160921	AR160921 Sequence
c	3	19	31.7	89959	8	AP004967	AP004967 Lotus cor
	4	19	31.7	200849	2	AC026895	AC026895 Homo sapi
c	5	18	30.0	59	6	AR160920	AR160920 Sequence
	6	18	30.0	201	11	BV203590	BV203590 sqnm21259
	7	18	30.0	201	11	BV203591	BV203591 sqnm21259
	8	18	30.0	201	11	BV203592	BV203592 sqnm21259
	9	18	30.0	432	6	AX397069	AX397069 Sequence
	10	18	30.0	1428	6	BD094070	BD094070 Shear str
	11	18	30.0	1428	9	HUMSF2P33	M69040 Human SF2p3
	12	18	30.0	1618	5	BC042354	BC042354 Xenopus l
	13	18	30.0	1717	9	HUMASF	M72709 Human alter
	14	18	30.0	1860	5	BC056752	BC056752 Danio rer
	15	18	30.0	2002	5	BC055511	BC055511 Danio rer
	16	18	30.0	2055	5	BC075558	BC075558 Xenopus t
	17	18	30.0	2059	9	BC033785	BC033785 Homo sapi
	18	18	30.0	2369	5	BC046679	BC046679 Xenopus l
c	19	18	30.0	2583	9	AB000463	AB000463 Homo sapi
	20	18	30.0	2708	9	BC010264	BC010264 Homo sapi
	21	18	30.0	2731	5	BC076945	BC076945 Xenopus t
	22	18	30.0	2765	6	BD189938	BD189938 Tumor ant
	23	18	30.0	2765	9	AB062124	AB062124 Homo sapi
	24	18	30.0	2900	5	BC066682	BC066682 Danio rer
	25	18	30.0	22970	9	HSL247F6	Z68279 Human DNA s
c	26	18	30.0	35240	2	AC040987	AC040987 Homo sapi
c	27	18	30.0	69023	2	AC087445	AC087445 Homo sapi
c	28	18	30.0	110000	2	BX927305_0	BX927305 Danio rer
c	29	18	30.0	119974	9	AC018763	AC018763 Homo sapi
	30	18	30.0	135299	9	AC091857	AC091857 Homo sapi
	31	18	30.0	142801	2	AC079206	AC079206 Homo sapi
	32	18	30.0	149969	9	HSJ323A24	AL121750 Human DNA
	33	18	30.0	154462	2	AC036136	AC036136 Homo sapi
	34	18	30.0	160111	2	AP000780	AP000780 Homo sapi
	35	18	30.0	160415	2	AC140961	AC140961 Papio anu
	36	18	30.0	164404	2	AC027149	AC027149 Homo sapi
	37	18	30.0	165726	2	AC143663	AC143663 Macaca mu
c	38	18	30.0	168651	9	AC009474	AC009474 Homo sapi
c	39	18	30.0	169403	9	AL445929	AL445929 Human DNA
	40	18	30.0	176310	2	BX897686	BX897686 Danio rer
c	41	18	30.0	176879	5	AL929224	AL929224 Zebrafish
c	42	18	30.0	177314	9	AC022209	AC022209 Homo sapi
c	43	18	30.0	181275	2	BX511308	BX511308 Danio rer
c	44	18	30.0	181494	9	AC015813	AC015813 Homo sapi
c	45	18	30.0	196476	2	AC055890	AC055890 Homo sapi

ALIGNMENTS

RESULT 1
AR160916
LOCUS AR160916 60 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 7 from patent US 6255281.
ACCESSION AR160916
VERSION AR160916.1 GI:16225981
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 60)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
TITLE Use of recombinant human uteroglobin in treatment of inflammatory and fibrotic conditions
JOURNAL Patent: US 6255281-A 7 03-JUL-2001;
FEATURES Location/Qualifiers
source 1. .60
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 60; DB 6; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.4e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60

RESULT 2
AR160921/c
LOCUS AR160921 60 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 12 from patent US 6255281.
ACCESSION AR160921
VERSION AR160921.1 GI:16225996
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 60)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.

```
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 12 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             61.7%;  Score 37;  DB 6;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-09;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 3
AP004967/c
LOCUS       AP004967               89959 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 1,
            clone:LjT27L02, TM0144, complete sequence.
ACCESSION   AP004967
VERSION     AP004967.1  GI:21907985
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence
            Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb
            Regions of the Genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89959)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:yn@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .89959
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT27L02"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0144~synonym: Lotus japonicus"
ORIGIN

  Query Match             31.7%;  Score 19;  DB 8;  Length 89959;
  Best Local Similarity   100.0%;  Pred. No. 6;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 AAGCAGCTATGGAACTGTT 26
              |||||||||||||||||||
Db      87202 AAGCAGCTATGGAACTGTT 87184


RESULT 4
AC026895
LOCUS       AC026895               200849 bp    DNA     linear   HTG 20-APR-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-396J2 map 2, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC026895
VERSION     AC026895.2  GI:7596820
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200849)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-396J2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200849)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 19, 2000 this sequence version replaced gi:7328755.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6461
                Center clone name: 396_J_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 188437 bases at least Q40
                Consensus quality: 194805 bases at least Q30
                Consensus quality: 197132 bases at least Q20
                Insert size: 210000; agarose-fp
                Insert size: 198749; sum-of-contigs
                Quality coverage: 4.6 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1067: contig of 1067 bp in length
            *     1068     1167: gap of 100 bp
            *     1168     2856: contig of 1689 bp in length
            *     2857     2956: gap of 100 bp
            *     2957     5983: contig of 3027 bp in length
            *     5984     6083: gap of 100 bp
```

```
            *     6084     7783: contig of 1700 bp in length
            *     7784     7883: gap of 100 bp
            *     7884    10615: contig of 2732 bp in length
            *    10616    10715: gap of 100 bp
            *    10716    12692: contig of 1977 bp in length
            *    12693    12792: gap of 100 bp
            *    12793    14959: contig of 2167 bp in length
            *    14960    15059: gap of 100 bp
            *    15060    19764: contig of 4705 bp in length
            *    19765    19864: gap of 100 bp
            *    19865    22666: contig of 2802 bp in length
            *    22667    22766: gap of 100 bp
            *    22767    27132: contig of 4366 bp in length
            *    27133    27232: gap of 100 bp
            *    27233    31434: contig of 4202 bp in length
            *    31435    31534: gap of 100 bp
            *    31535    39440: contig of 7906 bp in length
            *    39441    39540: gap of 100 bp
            *    39541    46513: contig of 6973 bp in length
            *    46514    46613: gap of 100 bp
            *    46614    55509: contig of 8896 bp in length
            *    55510    55609: gap of 100 bp
            *    55610    66256: contig of 10647 bp in length
            *    66257    66356: gap of 100 bp
            *    66357    77020: contig of 10664 bp in length
            *    77021    77120: gap of 100 bp
            *    77121    88050: contig of 10930 bp in length
            *    88051    88150: gap of 100 bp
            *    88151    98020: contig of 9870 bp in length
            *    98021    98120: gap of 100 bp
            *    98121   108280: contig of 10160 bp in length
            *   108281   108380: gap of 100 bp
            *   108381   126228: contig of 17848 bp in length
            *   126229   126328: gap of 100 bp
            *   126329   165233: contig of 38905 bp in length
            *   165234   165333: gap of 100 bp
            *   165334   200849: contig of 35516 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200849
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-396J2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1067
                     /note="assembly_fragment"
     misc_feature    1168. .2856
                     /note="assembly_fragment"
     misc_feature    2957. .5983
                     /note="assembly_fragment"
     misc_feature    6084. .7783
                     /note="assembly_fragment"
     misc_feature    7884. .10615
                     /note="assembly_fragment"
     misc_feature    10716. .12692
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    12793. .14959
                     /note="assembly_fragment"
     misc_feature    15060. .19764
                     /note="assembly_fragment"
     misc_feature    19865. .22666
                     /note="assembly_fragment"
     misc_feature    22767. .27132
                     /note="assembly_fragment"
     misc_feature    27233. .31434
                     /note="assembly_fragment"
     misc_feature    31535. .39440
                     /note="assembly_fragment"
     misc_feature    39541. .46513
                     /note="assembly_fragment"
     misc_feature    46614. .55509
                     /note="assembly_fragment"
     misc_feature    55610. .66256
                     /note="assembly_fragment"
     misc_feature    66357. .77020
                     /note="assembly_fragment"
     misc_feature    77121. .88050
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    88151. .98020
                     /note="assembly_fragment"
     misc_feature    98121. .108280
                     /note="assembly_fragment"
     misc_feature    108381. .126228
                     /note="assembly_fragment"
     misc_feature    126329. .165233
                     /note="assembly_fragment"
     misc_feature    165334. .200849
                     /note="assembly_fragment"
ORIGIN

  Query Match             31.7%;  Score 19;  DB 2;  Length 200849;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          9 AGCAGCTATGGAACTGTTC 27
              |||||||||||||||||||
Db     114115 AGCAGCTATGGAACTGTTC 114133


RESULT 5
AR160920/c
LOCUS       AR160920                 59 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 11 from patent US 6255281.
ACCESSION   AR160920
VERSION     AR160920.1  GI:16225993
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 59)
  AUTHORS   Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 11 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .59
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             30.0%;  Score 18;  DB 6;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42


RESULT 6
BV203590
LOCUS       BV203590                201 bp    DNA     linear   STS 10-JUN-2004
DEFINITION  sqnm212590 Human DNA (Sequenom) Homo sapiens STS genomic, sequence
            tagged site.
ACCESSION   BV203590
VERSION     BV203590.1  GI:48173010
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Nelson,R.M., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
            Cantor,C.R. and Braun,A.
  TITLE     Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
            Regions
  JOURNAL   Genome Res. (2004) In press
COMMENT
            Contact: Andreas Braun
            Pharmaceuticals division
            Sequenom, Inc.
            3595 John Hopkins Court, San Diego, CA 92121, USA
            Tel: 18582029018
            Fax: 18582029020
            Email: abraun@sequenom.com
            Primer A: No primer sequence submitted
            Primer B: No primer sequence submitted
            STS size: 201.
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human DNA (Sequenom)"
     STS             <1. .>201
ORIGIN

  Query Match             30.0%;  Score 18;  DB 11;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db         130 ACATGCGTGAAGCAGGTG 147


RESULT 7
BV203591
LOCUS       BV203591                201 bp    DNA     linear   STS 10-JUN-2004
DEFINITION  sqnm212591 Human DNA (Sequenom) Homo sapiens STS genomic, sequence
            tagged site.
ACCESSION   BV203591
VERSION     BV203591.1  GI:48173011
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Nelson,R.M., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
            Cantor,C.R. and Braun,A.
  TITLE     Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
            Regions
  JOURNAL   Genome Res. (2004) In press
COMMENT
            Contact: Andreas Braun
            Pharmaceuticals division
            Sequenom, Inc.
            3595 John Hopkins Court, San Diego, CA 92121, USA
            Tel: 18582029018
            Fax: 18582029020
            Email: abraun@sequenom.com
            Primer A: No primer sequence submitted
            Primer B: No primer sequence submitted
            STS size: 201.
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human DNA (Sequenom)"
     STS             <1. .>201
ORIGIN

  Query Match             30.0%;  Score 18;  DB 11;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db         121 ACATGCGTGAAGCAGGTG 138


RESULT 8
BV203592
LOCUS       BV203592                201 bp    DNA     linear   STS 10-JUN-2004
DEFINITION  sqnm212593 Human DNA (Sequenom) Homo sapiens STS genomic, sequence
            tagged site.
ACCESSION   BV203592
VERSION     BV203592.1  GI:48173012
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Nelson,R.M., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
            Cantor,C.R. and Braun,A.
  TITLE     Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
            Regions
  JOURNAL   Genome Res. (2004) In press
COMMENT
            Contact: Andreas Braun
            Pharmaceuticals division
            Sequenom, Inc.
            3595 John Hopkins Court, San Diego, CA 92121, USA
            Tel: 18582029018
            Fax: 18582029020
            Email: abraun@sequenom.com
            Primer A: No primer sequence submitted
            Primer B: No primer sequence submitted
            STS size: 201.
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human DNA (Sequenom)"
     STS             <1. .>201
ORIGIN

  Query Match             30.0%;  Score 18;  DB 11;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db         112 ACATGCGTGAAGCAGGTG 129


RESULT 9
AX397069
LOCUS       AX397069                432 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1284 from Patent WO0212328.
ACCESSION   AX397069
VERSION     AX397069.1  GI:21067816
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   King,G.E., Meagher,M.J., Xu,J. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0212328-A 1284 14-FEB-2002;
```

```
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             30.0%;  Score 18;  DB 6;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db           9 ACATGCGTGAAGCAGGTG 26


RESULT 10
BD094070
LOCUS       BD094070                1428 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Shear stress-responsive DNAs.
ACCESSION   BD094070
VERSION     BD094070.1  GI:22639658
KEYWORDS    WO 0125427-A/31.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1428)
  AUTHORS   Nojima,H., Yoshisue,H., Obayashi,M., Ota,T., Kawabata,A.,
            Sakurada,K., Kuga,T., Sekine,S., Nakamura,Y. and Sugan,S.
  TITLE     Shear stress-responsive DNAs
  JOURNAL   Patent: WO 0125427-A 31 12-APR-2001;
            KYOWA HAKKO KOGYO CO LTD,HIROSHI NOJIMA,HAJIME YOSHISUE, MASAYA
            OBAYASHI, TOSHIO OTA,AYAKO KAWABATA,KAZUHIRO SAKURADA,TETSURO KUGA,
            SUSUMU SEKINE, YUSUKE NAKAMURA,SUMIO SUGANO
COMMENT     OS   Homo sapiens (human)
            PN   WO 0125427-A/31
            PD   12-APR-2001
            PF   02-OCT-2000 WO 2000JP006840
            PR   01-OCT-1999 JP 99P   280976
            PI   HIROSHI NOJIMA,HAJIME YOSHISUE,MASAYA OBAYASHI,TOSHIO OTA, PI
             AYAKO KAWABATA,
            PI   KAZUHIRO SAKURADA,TETSURO KUGA,SUSUMU SEKINE,YUSUKE NAKAMURA,
            PI   SUMIO SUGANO
            PC   C12N15/12,C07K14/435,C07K16/18,C12P21/02,C12Q1/68,A61K38/00,
            PC   A61K39/395,
            PC   A61K48/00,A61P9/10,G01N33/50,G01N33/53
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (125). .(868).
FEATURES             Location/Qualifiers
     source          1. .1428
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             30.0%;  Score 18;  DB 6;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db         543 ACATGCGTGAAGCAGGTG 560


RESULT 11
HUMSF2P33
LOCUS       HUMSF2P33               1428 bp    mRNA    linear   PRI 24-NOV-1993
DEFINITION  Human SF2p33 mRNA, complete cds.
ACCESSION   M69040
VERSION     M69040.1  GI:338046
KEYWORDS    SF2p33.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1428)
  AUTHORS   Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G.
  TITLE     Functional expression of cloned human splicing factor SF2: homology
            to RNA-binding proteins, U1 70K, and Drosophila splicing regulators
  JOURNAL   Cell 66 (2), 383-394 (1991)
  MEDLINE   91309150
   PUBMED   1830244
COMMENT     Original source text: Homo sapiens cDNA to mRNA.
            Comments: SF2p33 is an essential pre-mRNA splicing factor, which
            can also affect alternative 5' splice site selection in vitro by
            stimulating the use of proximal 5' splice sites.
            (Krainer, A.R., Conway, G.C., and Kozak, D. (1990) Purification and
            Characterization of SF2, a Human Pre-mRNA Splicing Factor. Genes
            Dev. 4, 1158-1171; Krainer, A.R., Conway, G.C., and Kozak, D.
            (1990) The Essential Pre-mRNA Splicing Factor SF2 Influences 5'
            Splice Site Selection by Activating Proximal Sites. Cell 62,
            35-42). This factor is also known as ASF (Ge, H., and
            Manley, J.L. (1990) A protein Factor, ASF, Controls Alternative
            Splicing of SV40 Early Pre-mRNA In Vitro. Cell 62, 25-34). SF2 p33
            consists of two polypeptides of apparent molecular weight 33 kd,
            approximately, although their predicted molecular weight is 27,744
            daltons. The two forms appear to differ by the extent of
            post-translational modification, which includes phosphorylation.
            SF2p33 binds RNA and promotes the annealing of complementary RNAs.
            It is required for assembly of pre-spliceosome complexes. The
            N-termini of the HeLa polypeptides are blocked.
FEATURES             Location/Qualifiers
     source          1. .1428
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="female"
     5'UTR           <1. .124
     CDS             125. .871
                     /function="pre-mRNA splicing factor"
                     /codon_start=1
                     /product="SF2p33"
                     /protein_id="AAA03476.1"
                     /db_xref="GI:338047"
                     /translation="MSGGGVIRGPAGNNDCRIYVGNLPPDIRTKDIEDVFYKYGAIRD
                     IDLKNRRGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPRSGRGTGRGGGGG
                     GGGGAPRGRYGPPSRRSENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVV
                     EFVRKEDMTYAVRKLDNTKFRSHEGETAYIRVKVDGPRSPSYGRSRSRSRSRSRSRSR
                     SNSRSRSYSPRRSRGSPRYSPRHSRSRSRT"
     3'UTR           872. .1428
ORIGIN

  Query Match             30.0%;  Score 18;  DB 9;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          41 ACATGCGTGAAGCAGGTG 58
               ||||||||||||||||||
Db         543 ACATGCGTGAAGCAGGTG 560


RESULT 12
BC042354
LOCUS       BC042354                1618 bp    mRNA    linear   VRT 03-FEB-2004
DEFINITION  Xenopus laevis similar to splicing factor, arginine/serine-rich 1
            (splicing factor 2, alternate splicing factor), mRNA (cDNA clone
            IMAGE:4681465), partial cds.
ACCESSION   BC042354
VERSION     BC042354.1  GI:27503246
KEYWORDS    .
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 1618)
  AUTHORS   Klein,S.L., Strausberg,R.L., Wagner,L., Pontius,J., Clifton,S.W.
            and Richardson,P.
  TITLE     Genetic and genomic tools for Xenopus research: The NIH Xenopus
            initiative
  JOURNAL   Dev. Dyn. 225 (4), 384-391 (2002)
  PUBMED    12454917
REFERENCE   2  (bases 1 to 1618)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED    12477932
REFERENCE   3  (bases 1 to 1618)
  AUTHORS   Klein,S. and Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2003) National Institutes of Health, Xenopus Gene
            Collection (XGC), National Institute of Child Health and Human
            Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
            20892-7510, USA
  REMARK    NIH-MGC Project
COMMENT     Contact: XGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Igor Dawid
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 94 Row: b Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .1618
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4681465"
                     /tissue_type="Embryo, stage 31/32, Xenopus"
                     /clone_lib="NICHD_XGC_Emb4"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            <1. .1618
                     /gene="Sfrs1"
                     /db_xref="LocusID:399226"
     CDS             <1. .853
                     /gene="Sfrs1"
                     /codon_start=2
                     /product="Sfrs1 protein"
                     /protein_id="AAH42354.1"
                     /db_xref="GI:27503247"
                     /db_xref="LocusID:399226"
                     /translation="ASIALFLYKPSAFKAIKLDTSAGGVIRGPAGNNDCRIYVGNLPP
                     DIRTKDIEDVFYKYGAIRDIDLKNRRGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYR
                     LRVEFPRSGRGAGGRGGGGGGGGSGGGGGGGGGGGGGGGAPRGRYGPPSRRSEYRVVV
                     SGLPPSGSWQDLKDHMREAGDVCYADVFRDGTGVVEFVRKEDMTYAVRKLDNTKFRSH
                     EGETAYIRVKVDGPRSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRH
                     SRSRSRT"
     misc_feature    107. .271
                     /gene="Sfrs1"
                     /note="RRM; Region: RNA recognition motif"
                     /db_xref="CDD:smart00360"
     misc_feature    476. .658
                     /gene="Sfrs1"
                     /note="RRM_1; Region: RNA recognition motif. (a.k.a. RRM,
                     RBD, or RNP domain). The RRM motif is probably diagnostic
                     of an RNA binding protein. RRMs are found in a variety of
                     RNA binding proteins, including various hnRNP proteins,
                     proteins implicated in regulation of alternative splicing,
                     and protein components of snRNPs. The motif also appears
                     in a few single stranded DNA binding proteins. The RRM
                     structure consists of four strands and two helices
                     arranged in an alpha/beta sandwich, with a third helix
                     present during RNA binding in some cases The C-terminal
                     beta strand (4th strand) and final helix are hard to align
                     and have been omitted in the SEED alignment The LA
                     proteins have a N terminus rrm which is included in the
                     seed. There is a second region towards the C terminus that
                     has some features of a rrm but does not appear to have the
                     important structural core of a rrm. The LA proteins are
                     one of the main autoantigens in Systemic lupus
                     erythematosus (SLE), an autoimmune disease"
                     /db_xref="CDD:pfam00076"
ORIGIN

  Query Match           30.0%;  Score 18;  DB 5;  Length 1618;
  Best Local Similarity  100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        525 ACATGCGTGAAGCAGGTG 542


RESULT 13
HUMASF
LOCUS       HUMASF                  1717 bp    mRNA    linear   PRI 27-APR-1993
DEFINITION  Human alternative splicing factor mRNA, complete cds.
ACCESSION   M72709
VERSION     M72709.1  GI:179073
KEYWORDS    alternative splicing factor.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1717)
  AUTHORS   Ge,H., Zuo,P. and Manley,J.L.
  TITLE     Primary structure of the human splicing factor ASF reveals
            similarities with Drosophila regulators
  JOURNAL   Cell 66 (2), 373-382 (1991)
  MEDLINE   91309149
  PUBMED    1855257
COMMENT     Original source text: Homo sapiens cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1717
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
     mRNA            join(1. .648,831. .1717)
                     /note="alternative"
```

```
     mRNA            join(1. .648,848. .1717)
                     /note="alternative"
     exon            1. .648
                     /number=1
     5'UTR           1. .96
     CDS             join(97. .648,831. .1157)
                     /note="alternative"
                     /number=2
                     /codon_start=1
                     /protein_id="AAA35564.1"
                     /db_xref="GI:179074"
                     /translation="MSGGGVIRGPAGNNDCRIYVGNLPPDIRTKDIEDVFYKYGAIRD
                     IDLKNRRGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPRSGRGTGRGGGGG
                     GGGGAPRGRYGPPSRRSENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVV
                     EFVRKEDMTYAVRKLDNTKFRSHEFCLSNREKLPTSGLKLMGPEVQVMEDLDLEAVVV
                     AEAVAEATAGVAVTPQGEAEDHHAILPVIADLALVHKMIGDTFCRTHVVYSFPLFSTI
                     FSFFNSNCFVQNGLKC"
     CDS             join(97. .648,848. .1042)
                     /note="alternative"
                     /number=1
                     /codon_start=1
                     /protein_id="AAA35565.1"
                     /db_xref="GI:179075"
                     /translation="MSGGGVIRGPAGNNDCRIYVGNLPPDIRTKDIEDVFYKYGAIRD
                     IDLKNRRGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPRSGRGTGRGGGGG
                     GGGGAPRGRYGPPSRRSENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVV
                     EFVRKEDMTYAVRKLDNTKFRSHEGETAYIRVKVDGPRSPSYGRSRSRSRSRSRSRSR
                     SNSRSRSYSPRRSRGSPRYSPRHSRSRSRT"
     intron          649. .847
                     /number=1
                     /label=1a
     intron          649. .830
                     /number=1
                     /label=1b
     exon            831. .1717
                     /number=2
                     /label=2b
     exon            848. .1717
                     /number=2
                     /label=2a
     3'UTR           1043. .1717
ORIGIN

  Query Match           30.0%;  Score 18;  DB 9;  Length 1717;
  Best Local Similarity  100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        515 ACATGCGTGAAGCAGGTG 532


RESULT 14
BC056752
LOCUS       BC056752               1860 bp    mRNA    linear   VRT 25-MAY-2004
DEFINITION  Danio rerio cDNA clone MGC:65898 IMAGE:6801570, complete cds.
ACCESSION   BC056752
VERSION     BC056752.1  GI:34785173
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 1860)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
    TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1860)
  AUTHORS   Strausberg,R.
    TITLE   Direct Submission
  JOURNAL   Submitted (25-AUG-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
   REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Chi-Bin Chien
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 121 Row: p Column: 12
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES                Location/Qualifiers
     source             1. .1860
                        /organism="Danio rerio"
                        /mol_type="mRNA"
                        /db_xref="taxon:7955"
                        /clone="MGC:65898 IMAGE:6801570"
                        /tissue_type="Embryo, zebrafish whole embryo 72 hpf"
                        /clone_lib="NCI_CGAP_ZEmb3"
                        /lab_host="DH10B"
                        /note="Vector: pCMV-SPORT6.1"
     CDS                196. .762
                        /codon_start=1
                        /product="Unknown (protein for MGC:65898)"
                        /protein_id="AAH56752.1"
                        /db_xref="GI:34785174"
                        /translation="MSGGVIRGPAGNNDCRIYVGNLPPDIRTKDVEDVFYKYGAIRDI
                        DLKNRRGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPRSGRGGGRGGGGGG
                        GVGAPRGRYGPPSRRSEYRVIVSGLPPSGSWQDLKDHMREAGDVCYADVFRDGTGVVE
                        FVRKEDMTYAVRKLDNTKFRSHEVGVSL"
ORIGIN

  Query Match           30.0%;  Score 18;  DB 5;  Length 1860;
  Best Local Similarity  100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
```

```
                    ||||||||||||||||||
Db        611 ACATGCGTGAAGCAGGTG 628


RESULT 15
BC055511
LOCUS       BC055511                2002 bp    mRNA    linear   VRT 30-JUN-2004
DEFINITION  Danio rerio zgc:66146, mRNA (cDNA clone MGC:66146 IMAGE:5410436),
            complete cds.
ACCESSION   BC055511
VERSION     BC055511.1  GI:33416586
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 2002)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2002)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Sumio Sugano
            cDNA Library Preparation: Dr. Sumio Sugano
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 122 Row: f Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2002
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:66146 IMAGE:5410436"
                     /tissue_type="Whole body, adult male"
                     /clone_lib="Sugano SJD adult male"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3"
     gene            1. .2002
                     /gene="zgc:66146"
                     /db_xref="LocusID:406288"
     CDS             80. .817
                     /gene="zgc:66146"
                     /codon_start=1
                     /product="zgc:66146"
                     /protein_id="AAH55511.1"
                     /db_xref="GI:33416587"
                     /db_xref="LocusID:406288"
                     /translation="MSGGVIRGPAGSNDCRIYVGNLPPDIRTKDVEDVFYKYGAIRDI
                     DLKNRRGGPPFAFVEFEDPRDAEDAVYARDGYDYDGYRLRVEFPRSGRGMGRGGFGGG
                     GGGGGGGGGGGGGGAPRGRYGPPSRRSEYRVIVSGLPPSGSWQDLKDHMREAGDVCYA
                     DVFRDGTGVVEFVRKEDMTYAVRKLDNTKFRSHEGETAYIRVKVDGPRSPSYGRSRSR
                     SRSRSRSRSRSNNRSRSYSPRRSRGSP"
ORIGIN

  Query Match           30.0%;  Score 18;  DB 5;  Length 2002;
  Best Local Similarity  100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        528 ACATGCGTGAAGCAGGTG 545


Search completed: December 22, 2004, 23:36:31
Job time : 854.688 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:14:28 ; Search time 211.765 Seconds
(without alignments)
1487.336 Million cell updates/sec

Title: US-09-898-616A-2
Perfect score: 60
Sequence: 1 agctacgaagcagctatgga.....acatgcgtgaagcaggtgct 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60	100.0	60	9	ABZ58371	Abz58371 Human ute
	2	60	100.0	60	12	ADL27627	Adl27627 Recombina
c	3	37	61.7	60	9	ABZ58376	Abz58376 Human ute
c	4	37	61.7	60	9	ABZ58377	Abz58377 Human ute
c	5	37	61.7	60	12	ADL27633	Adl27633 Recombina
c	6	37	61.7	60	12	ADL27632	Adl27632 Recombina
c	7	18	30.0	59	9	ABZ58375	Abz58375 Human ute
c	8	18	30.0	59	12	ADL27631	Adl27631 Recombina
	9	18	30.0	432	6	ABK45733	Abk45733 cDNA enco
	10	18	30.0	489	9	ACH35932	Ach35932 Human end
	11	18	30.0	1428	4	AAH02904	Aah02904 Human she
	12	18	30.0	1428	10	AAL60055	Aal60055 Human Pc0
	13	18	30.0	1717	10	ADH28741	Adh28741 Human chr
	14	18	30.0	1926	12	ADP07658	Adp07658 Human sec
	15	18	30.0	2765	10	ADD22452	Add22452 HLA-B46 T
	16	18	30.0	2765	10	ADI15992	Adi15992 Human PP
	17	18	30.0	3252	12	ADQ22406	Adq22406 Human sof
	18	18	30.0	3299	10	ADI22703	Adi22703 Human liv
	19	18	30.0	22970	10	ADK70082	Adk70082 Mutant hu
	20	18	30.0	22970	10	ADK70081	Adk70081 Wild type
	21	17	28.3	29	6	AAS18797	Aas18797 PCR prime
	22	17	28.3	1978	6	ABN59992	Abn59992 Novel hum
	23	16	26.7	65	6	ABN56135	Abn56135 Mouse spl
c	24	16	26.7	300	2	AAZ14356	Aaz14356 Human gen
	25	16	26.7	351	12	ADL85790	Adl85790 DNA up-re
	26	16	26.7	351	12	ADL85789	Adl85789 DNA up-re
	27	16	26.7	373	8	ABX50810	Abx50810 Bovine ES
	28	16	26.7	624	9	ADA30641	Ada30641 DNA encod
c	29	16	26.7	732	2	AAZ17762	Aaz17762 Human gen
c	30	16	26.7	802	2	AAZ16289	Aaz16289 Human gen
c	31	16	26.7	830	11	ADI31178	Adi31178 Human cDN
c	32	16	26.7	846	3	AAC49282	Aac49282 Arabidops
	33	16	26.7	869	6	ABK13152	Abk13152 Transcrip
c	34	16	26.7	1019	4	AAS59766	Aas59766 Propionib
c	35	16	26.7	1019	8	ACF64695	Acf64695 Propionib
	36	16	26.7	1820	10	ADD18824	Add18824 Human dis
c	37	16	26.7	1919	11	ADM03518	Adm03518 Human cDN
	38	16	26.7	1959	11	ABD09379	Abd09379 Pseudomon
c	39	16	26.7	2010	11	ABD09565	Abd09565 Pseudomon
c	40	16	26.7	2402	4	AAK81667	Aak81667 Human imm
	41	16	26.7	3602	8	ACF12861	Acf12861 Human cer
	42	16	26.7	3803	2	AAX76396	Aax76396 Human sec
	43	16	26.7	3803	10	ADC38821	Adc38821 Human cDN
	44	16	26.7	5333	10	ADD47032	Add47032 Human gen
	45	16	26.7	5333	10	ADD47860	Add47860 Human gen

ALIGNMENTS

RESULT 1
ABZ58371
ID ABZ58371 standard; DNA; 60 BP.
XX
AC ABZ58371;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 2.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Claim 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 2, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and

CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 15 A; 15 C; 18 G; 12 T; 0 U; 0 Other;

Query Match 100.0%; Score 60; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.2e-23;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60

RESULT 2
ADL27627
ID ADL27627 standard; DNA; 60 BP.
XX
AC ADL27627;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, coding oligonucleotide #2.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Claim 1; SEQ ID NO 2; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a coding strand oligonucleotide used to
CC construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 15 A; 15 C; 18 G; 12 T; 0 U; 0 Other;

Query Match 100.0%; Score 60; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.2e-23;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60

RESULT 3
ABZ58376/c
ID ABZ58376 standard; DNA; 60 BP.
XX
AC ABZ58376;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 7.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 7, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the

CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 61.7%; Score 37; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
||||||||||||||||||||||||||||||||||||||
Db 37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1

RESULT 4
ABZ58377/c
ID ABZ58377 standard; DNA; 60 BP.
XX
AC ABZ58377;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 8.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 8, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 61.7%; Score 37; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
||||||||||||||||||||||||||||||||||||||
Db 37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1

RESULT 5
ADL27633/c
ID ADL27633 standard; DNA; 60 BP.
XX
AC ADL27633;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #4.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 8; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a

CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 61.7%; Score 37; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
||||||||||||||||||||||||||||||||||||||
Db 37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1

RESULT 6
ADL27632/c
ID ADL27632 standard; DNA; 60 BP.
XX
AC ADL27632;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #3.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 7; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 12 A; 14 C; 22 G; 12 T; 0 U; 0 Other;

Query Match 61.7%; Score 37; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
||||||||||||||||||||||||||||||||||||||
Db 37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1

RESULT 7
ABZ58375/c
ID ABZ58375 standard; DNA; 59 BP.
XX
AC ABZ58375;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 6.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX

PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 6, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 59 BP; 9 A; 17 C; 15 G; 18 T; 0 U; 0 Other;

Query Match 30.0%; Score 18; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AGGACATGCGTGAAGCAG 55
||||||||||||||||||
Db 59 AGGACATGCGTGAAGCAG 42

RESULT 8
ADL27631/c
ID ADL27631 standard; DNA; 59 BP.
XX
AC ADL27631;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #2.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 6; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 59 BP; 9 A; 17 C; 15 G; 18 T; 0 U; 0 Other;

Query Match 30.0%; Score 18; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AGGACATGCGTGAAGCAG 55
||||||||||||||||||
Db 59 AGGACATGCGTGAAGCAG 42

RESULT 9
ABK45733
ID ABK45733 standard; cDNA; 432 BP.
XX
AC ABK45733;
XX
DT 05-JUN-2002 (first entry)
XX
DE cDNA encoding colon tumour protein, SEQ ID No 1284.
XX
KW Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212328-A2.
XX
PD 14-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US024218.
XX
PR 03-AUG-2000; 2000US-0223283P.
PR 28-MAR-2001; 2001US-0279763P.
PR 29-JUN-2001; 2001US-0302051P.
XX
PA (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secrist H;
XX
DR WPI; 2002-241739/29.
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,

PT for diagnosing, preventing, and treating colon cancer, and as markers for
PT the progression of cancer.
XX
PS Claim 1; SEQ ID NO 1284; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer. ABK44450-
CC ABK46237 represent coding sequences of human colon tumour proteins of the
CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
CC data for this patent did not form part of the printed specification but
CC was supplied by the European Patent Office
XX
SQ Sequence 432 BP; 123 A; 85 C; 103 G; 121 T; 0 U; 0 Other;

Query Match 30.0%; Score 18; DB 6; Length 432;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 ACATGCGTGAAGCAGGTG 58
||||||||||||||||||
Db 9 ACATGCGTGAAGCAGGTG 26

RESULT 10
ACH35932
ID ACH35932 standard; cDNA; 489 BP.
XX
AC ACH35932;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human endothelial cell cDNA #4065.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 23144; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 489 BP; 130 A; 103 C; 134 G; 119 T; 0 U; 3 Other;

Query Match 30.0%; Score 18; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 ACATGCGTGAAGCAGGTG 58
||||||||||||||||||
Db 122 ACATGCGTGAAGCAGGTG 139

RESULT 11
AAH02904
ID AAH02904 standard; DNA; 1428 BP.
XX
AC AAH02904;
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 61.
XX
KW Human; shear stress-response protein; vascular disease; arteriosclerosis;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200125427-A1.
XX
PD 12-APR-2001.
XX
PF 02-OCT-2000; 2000WO-JP006840.
XX
PR 01-OCT-1999; 99JP-00280976.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (NOJI/) NOJIMA H.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI; 2001-266308/27.
DR P-PSDB; AAB90781.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis.
XX
PS Claim 20; Page 386-388; 678pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseasss caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension
XX
SQ Sequence 1428 BP; 341 A; 296 C; 400 G; 391 T; 0 U; 0 Other;

Query Match 30.0%; Score 18; DB 4; Length 1428;

```
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        543 ACATGCGTGAAGCAGGTG 560


RESULT 12
AAL60055
ID   AAL60055 standard; cDNA; 1428 BP.
XX
AC   AAL60055;
XX
DT   27-AUG-2003  (first entry)
XX
DE   Human Pc036-2 cDNA.
XX
KW   Human; differentially regulated protein; prevention; therapy; vaccine;
KW   prostate cancer; gene therapy; pre-mRNA splicing factor; Pc036-2;
KW   chromosome 17q21.3-q22; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             125. .871
FT                   /*tag= a
FT                   /product= "Human Pc036-2 protein"
XX
PN   WO2003040331-A2.
XX
PD   15-MAY-2003.
XX
PF   07-NOV-2002; 2002WO-US035563.
XX
PR   07-NOV-2001; 2001US-0331041P.
PR   07-NOV-2001; 2001US-0331042P.
PR   18-DEC-2001; 2001US-0340251P.
PR   07-JAN-2002; 2002US-0344791P.
XX
PA   (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI   Sun Z,  Li X,  Jay G,  Kovacs KF,  Fan W;
XX
DR   WPI; 2003-449451/42.
DR   P-PSDB; AAO29561.
XX
PT   New polynucleotide for diagnosing, staging, monitoring, prognosticating,
PT   preventing or treating, or determining the predisposition to, diseases or
PT   conditions such as prostate cancer, and for research or forensic science.
XX
PS   Claim 29; Page 127-128; 100pp; English.
XX
CC   The present invention relates to novel differentially regulated genes and
CC   polypeptides encoded by them. Sequences of the invention are useful in
CC   diagnosing, staging, monitoring, prognosticating, preventing, treating or
CC   determining the predisposition to diseases or conditions such as prostate
CC   cancer. They may be used as molecular markers, drug targets, vaccines, in
CC   gene therapy, research, clinical medicine or forensic science. The
CC   present sequence is a differentially regulated prostate cDNA, Pc036-2
CC   which codes for a pre-mRNA splicing factor. Pc036-2 gene is located on
CC   chromosome 17q21.3-q22
XX
SQ   Sequence 1428 BP; 341 A; 296 C; 400 G; 391 T; 0 U; 0 Other;

  Query Match            30.0%;  Score 18;  DB 10;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        543 ACATGCGTGAAGCAGGTG 560


RESULT 13
ADH28741
ID   ADH28741 standard; DNA; 1717 BP.
XX
AC   ADH28741;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human chronic myclogenous leukaemia (CML) gene marker #9.
XX
KW   ds; chronic phase chronic myelogenous leukaemia; CP-CML;
KW   blast crisis CML; BC-CML; human; chronic myclogenous leukaemia;
KW   gene marker.
XX
OS   Homo sapiens.
XX
PN   US2003104426-A1.
XX
PD   05-JUN-2003.
XX
PF   14-JUN-2002; 2002US-00171581.
XX
PR   18-JUN-2001; 2001US-0298914P.
XX
PA   (LINS/) LINSLEY P S.
PA   (MAOM/) MAO M.
PA   (DAIH/) DAI H.
PA   (HEYY/) HE Y.
PA   (RADI/) RADICH J P.
XX
PI   Linsley PS,  Mao M,  Dai H,  He Y,  Radich JP;
XX
DR   WPI; 2003-787046/74.
XX
PT   Classifying cell sample as chronic phase chronic myelogenous leukemia or
PT   blast crisis chronic myelogenous leukemia by detecting difference in
PT   expression of genes corresponding to the markers such as X15415, U89436.
XX
PS   Disclosure; SEQ ID NO 9; 31pp; English.
XX
CC   The invention relates to a method of classifying a cell sample as chronic
CC   phase chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-
CC   CML). The method is useful for classifying a sample as CP-CML or BC-CML.
CC   The present sequence represents a human chronic myclogenous leukaemia
CC   (CML) gene marker used to distinguish blast crisis CML from chronic phase
CC   CML.
XX
SQ   Sequence 1717 BP; 420 A; 343 C; 465 G; 489 T; 0 U; 0 Other;

  Query Match            30.0%;  Score 18;  DB 10;  Length 1717;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db        515 ACATGCGTGAAGCAGGTG 532


RESULT 14
ADP07658
ID   ADP07658 standard; DNA; 1926 BP.
XX
AC   ADP07658;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human secreted protein encoding DNA, seq id 141.
XX
KW   Cytostatic; antidiabetic; anorectic; gynaecological; antipsoriatic;
KW   dermatological; antiarteriosclerotic; antiasthmatic; neuroprotective;
KW   nootropic; antiparkinsonian; nephrotropic; human; secreted protein;
KW   diagnostic; pharmaceutical; cancer; lung; oesophageal; liver; diabetes;
```

KW obesity; metabolic disorder; cardiovascular disorder;
KW reproductive disorder; psoriasis; eczema; bronchitis; cystic fibrosis;
KW atherosclerosis; benign prostatic hyperplasia; asthma;
KW Alzheimer's disease; Parkinson's disease; renal disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004042000-A2.
XX
PD 21-MAY-2004.
XX
PF 16-MAY-2003; 2003WO-US015439.
XX
PR 17-MAY-2002; 2002US-0381592P.
PR 12-JUN-2002; 2002US-0388543P.
PR 08-AUG-2002; 2002US-0401757P.
PR 12-AUG-2002; 2002US-0402585P.
PR 13-AUG-2002; 2002US-0402799P.
PR 22-AUG-2002; 2002US-0404959P.
PR 04-OCT-2002; 2002US-0415902P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Olsen H, Baker KP, Fiscella M, Wei P;
PI Birse CE, Komatsoulis G, Choi GH, Moore PA, Gupta R, Shi Y;
XX
DR WPI; 2004-400658/37.
DR P-PSDB; ADP07840.
XX
PT New human secreted polypeptides and nucleic acid molecules for
PT diagnosing, preventing or treating disorders associated with the secreted
PT proteins, such as cancer, diabetes, obesity, cardiovascular disorders or
PT renal disorders.
XX
PS Claim 7; SEQ ID NO 141; 1157pp; English.
XX
CC The invention relates to a human secreted polypeptide for diagnosing,
CC preventing or treating disorders associated with the secreted proteins.
CC The polypeptides and nucleic acid molecules of the invention are useful
CC for preparing a diagnostic or pharmaceutical composition for diagnosing
CC or treating a medical condition. These may be used for diagnosing,
CC preventing or treating disorders related to the human secreted proteins,
CC such as cancer (e.g. lung, oesophageal or liver cancer), diabetes,
CC obesity, metabolic disorders, cardiovascular disorders, reproductive
CC disorders, psoriasis, eczema, bronchitis, cystic fibrosis,
CC atherosclerosis, benign prostatic hyperplasia, asthma, Alzheimer's
CC disease, Parkinson's disease or renal disorders. Sequences given in
CC records for ADP07528-ADP07709 represent human secreted protein encoding
CC DNA's of the invention.
XX
SQ Sequence 1926 BP; 479 A; 378 C; 568 G; 498 T; 0 U; 3 Other;

Query Match 30.0%; Score 18; DB 12; Length 1926;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 ACATGCGTGAAGCAGGTG 58
||||||||||||||||||
Db 997 ACATGCGTGAAGCAGGTG 1014

RESULT 15
ADD22452
ID ADD22452 standard; DNA; 2765 BP.
XX
AC ADD22452;
XX
DT 15-JAN-2004 (first entry)
XX
DE HLA-B46 T cell recognised tumour antigenic polypeptide, SEQ No 102.
XX
KW tumour antigenic peptide; cancer; vaccine; cytostatic; cytotoxic T cell;
KW colon; mouth; lung; prostatic; gynecological; human; gene; ds.
XX
OS Homo sapiens.
XX
PN JP2003111595-A.
XX
PD 15-APR-2003.
XX
PF 24-JUN-2002; 2002JP-00183603.
XX
PR 25-JUN-2001; 2001JP-00191974.
XX
PA (ITOY/) ITO Y.
XX
DR WPI; 2003-611129/58.
XX
PT Novel tumor antigenic peptide or polypeptide useful for inducing
PT cytotoxic T cells or for treating cancer such as colon, mouth, lung,
PT prostatic or gynecological cancer.
XX
PS Claim 10; SEQ ID NO 102; 98pp; Japanese.
XX
CC The invention relates to a novel tumour antigenic peptide or polypeptide
CC comprising a sequence selected from 99 sequences fully defined in the
CC specification. The tumour antigenic peptide or polypeptide comprises a
CC sequence selected from 99 sequences fully defined in the specification,
CC where the tumour antigenic peptide preferably has a sequence of Glu-Pro-
CC Pro-Leu-Ser-Gln-Glu-Thr-Phe, and the polypeptide preferably has a
CC sequence comprising 393 amino acids fully defined in the specification.
CC The invention further provides a cancer vaccine comprising a tumour
CC antigenic peptide or polypeptide, which has cytostatic activity. A tumour
CC antigenic peptide, polypeptide, its encoding polynucleotide, a
CC hybridising polynucleotide, a recombinant vector containing the
CC polynucleotide, a host transformed with the vector or an antibody are
CC useful for screening for compounds that interact with the tumour
CC antigenic peptide, the polypeptide or its encoding polynucleotide and
CC increases the expression of the tumour antigenic peptide, the polypeptide
CC or polynucleotide. The tumour antigenic peptide or the polypeptide is
CC useful for inducing cytotoxic T cells. The tumour antigenic peptide
CC vaccine is useful for treating cancer such as colon, mouth, lung,
CC prostatic or gynecological cancer. The invention also provides a
CC pharmaceutical composition useful for treating cancer. The tumour
CC antigenic peptide or the polypeptide is useful as an antigen to create
CC antibodies. This polynucleotide sequence represents the DNA encoding one
CC of the human tumour antigenic polypeptides of the invention.
XX
SQ Sequence 2765 BP; 759 A; 500 C; 652 G; 854 T; 0 U; 0 Other;

Query Match 30.0%; Score 18; DB 10; Length 2765;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 ACATGCGTGAAGCAGGTG 58
||||||||||||||||||
Db 509 ACATGCGTGAAGCAGGTG 526

Search completed: December 22, 2004, 22:44:12
Job time : 217.015 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 22:08:48 ; Search time 48.3258 Seconds
(without alignments)
882.496 Million cell updates/sec

Title: US-09-898-616A-2
Perfect score: 60
Sequence: 1 agctacgaagcagctatgga.....acatgcgtgaagcaggtgct 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		60	100.0	60	3	US-08-864-357F-7	Sequence 7, Appli
2	c	37	61.7	60	3	US-08-864-357F-12	Sequence 12, Appl
3	c	18	30.0	59	3	US-08-864-357F-11	Sequence 11, Appl
4		16	26.7	624	4	US-09-328-352-1928	Sequence 1928, Ap
5	c	16	26.7	830	4	US-09-023-655-504	Sequence 504, App
6		16	26.7	1959	4	US-09-252-991A-7983	Sequence 7983, Ap
7	c	16	26.7	2010	4	US-09-252-991A-8169	Sequence 8169, Ap
8		16	26.7	26700	1	US-08-472-217-1	Sequence 1, Appli
9		16	26.7	26700	2	US-08-488-199-5	Sequence 5, Appli
10		16	26.7	26700	3	US-08-760-534A-1	Sequence 1, Appli
11		16	26.7	26700	4	US-09-336-757-1	Sequence 1, Appli
12		15	25.0	540	4	US-09-270-767-3339	Sequence 3339, Ap
13		15	25.0	540	4	US-09-270-767-18621	Sequence 18621, A
14		15	25.0	1089	3	US-08-978-589A-1	Sequence 1, Appli
15		15	25.0	1089	3	US-09-336-601-2	Sequence 2, Appli
16		15	25.0	1089	4	US-09-219-120-1	Sequence 1, Appli
17	c	15	25.0	1651	4	US-09-976-594-598	Sequence 598, App
18	c	15	25.0	1913	2	US-08-588-258B-41	Sequence 41, Appl
19	c	15	25.0	1913	4	US-09-016-434-1078	Sequence 1078, Ap
20	c	15	25.0	1913	5	PCT-US96-08295-41	Sequence 41, Appl
21		15	25.0	1946	2	US-08-755-584-1	Sequence 1, Appli
22		15	25.0	1946	3	US-09-192-611-1	Sequence 1, Appli
23		15	25.0	1946	4	US-08-755-592A-5	Sequence 5, Appli
24		15	25.0	1946	4	US-09-617-923-1	Sequence 1, Appli
25		15	25.0	2432	4	US-09-799-451-5	Sequence 5, Appli
26		15	25.0	2513	4	US-09-799-451-6	Sequence 6, Appli
27		15	25.0	2900	1	US-08-034-650-9	Sequence 9, Appli
28		15	25.0	2900	1	US-08-449-015-9	Sequence 9, Appli
29		15	25.0	6151	4	US-09-799-451-528	Sequence 528, App
30	c	15	25.0	7194	4	US-09-601-326-76	Sequence 76, Appl
31	c	15	25.0	15420	4	US-09-601-326-54	Sequence 54, Appl
32	c	14	23.3	136	4	US-09-513-999C-19445	Sequence 19445, A
33		14	23.3	292	4	US-09-313-294A-957	Sequence 957, App
34		14	23.3	303	4	US-09-489-039A-4877	Sequence 4877, Ap
35		14	23.3	363	4	US-09-248-796A-2610	Sequence 2610, Ap
36	c	14	23.3	373	4	US-09-513-999C-9517	Sequence 9517, Ap
37		14	23.3	498	4	US-09-621-976-2381	Sequence 2381, Ap
38		14	23.3	575	4	US-09-270-767-27759	Sequence 27759, A
39		14	23.3	724	3	US-08-998-416-810	Sequence 810, App
40	c	14	23.3	798	4	US-09-489-039A-4894	Sequence 4894, Ap
41		14	23.3	836	2	US-08-698-805-7	Sequence 7, Appli
42	c	14	23.3	872	4	US-09-270-767-29234	Sequence 29234, A
43		14	23.3	879	4	US-09-620-312D-87	Sequence 87, Appl
44		14	23.3	1105	4	US-09-221-017B-108	Sequence 108, App
45		14	23.3	1176	4	US-09-270-767-12926	Sequence 12926, A

ALIGNMENTS

RESULT 1
US-08-864-357F-7
; Sequence 7, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 60
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-7

Query Match 100.0%; Score 60; DB 3; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.5e-24;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
```

RESULT 2
US-08-864-357F-12/c
; Sequence 12, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 60
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-12

```
  Query Match             61.7%;  Score 37;  DB 3;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 3
US-08-864-357F-11/c
; Sequence 11, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
;  APPLICANT:  Claragen, Inc. & NIH
;  TITLE OF INVENTION:   Use of Recombinant Human Uteroglobin in Treatment of Inflammatc
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE:   116142/2
;  CURRENT APPLICATION NUMBER: US/08/864,357F
;  CURRENT FILING DATE:  1997-05-28
;  NUMBER OF SEQ ID NOS:   22
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: artificial
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-08-864-357F-11

  Query Match             30.0%;  Score 18;  DB 3;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 0.62;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42


RESULT 4
US-09-328-352-1928
; Sequence 1928, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1928
;   LENGTH: 624
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-1928

  Query Match             26.7%;  Score 16;  DB 4;  Length 624;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         42 CATGCGTGAAGCAGGT 57
              ||||||||||||||||
Db        177 CATGCGTGAAGCAGGT 192


RESULT 5
US-09-023-655-504/c
; Sequence 504, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  504:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  830 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TESTNOT03
;      CLONE:  2006402
US-09-023-655-504

  Query Match             26.7%;  Score 16;  DB 4;  Length 830;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GCGTGAAGCAGGTGCT 60
              ||||||||||||||||
Db        195 GCGTGAAGCAGGTGCT 180


RESULT 6
US-09-252-991A-7983
; Sequence 7983, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 7983
;   LENGTH: 1959
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7983
```

```
  Query Match             26.7%;  Score 16;  DB 4;  Length 1959;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 GCGTGAAGCAGGTGCT 60
               ||||||||||||||||
Db        1913 GCGTGAAGCAGGTGCT 1928


RESULT 7
US-09-252-991A-8169/c
; Sequence 8169, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 8169
;   LENGTH: 2010
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8169

  Query Match             26.7%;  Score 16;  DB 4;  Length 2010;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          45 GCGTGAAGCAGGTGCT 60
               ||||||||||||||||
Db         134 GCGTGAAGCAGGTGCT 119


RESULT 8
US-08-472-217-1
; Sequence 1, Application US/08472217
; Patent No. 5726058
;  GENERAL INFORMATION:
;    APPLICANT:  Alanen-Kurki, Leena
;    APPLICANT:  Auvinen, Petri
;    APPLICANT:  Jaakkola, Panu
;    APPLICANT:  Jalkanen, Markku
;    APPLICANT:  Lepp , Sirpa
;    APPLICANT:  Mali, Markku
;    APPLICANT:  Vihinen, Tapani
;    APPLICANT:  W rri, Anni
;    TITLE OF INVENTION:  Syndecan Stimulation Of Cellular
;    TITLE OF INVENTION:  Differentiation
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,217
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,186
;      FILING DATE:  07-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/988,427
;      FILING DATE:  01-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cimbala, Michele A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1102.0050003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4378..4443, 22026..22106, 23001..23483,
;      LOCATION:  23905..24039, 24251..24418)
US-08-472-217-1

  Query Match             26.7%;  Score 16;  DB 1;  Length 26700;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          29 CTCCGGACCAGGACAT 44
               ||||||||||||||||
Db        3800 CTCCGGACCAGGACAT 3815


RESULT 9
US-08-488-199-5
; Sequence 5, Application US/08488199
; Patent No. 5851993
;  GENERAL INFORMATION:
;    APPLICANT:  Jalkanen, Markku
;    APPLICANT:  Mali, Markku
;    TITLE OF INVENTION:  Suppression of Tumor Cell Growth By
;    TITLE OF INVENTION:  Syndecan-1 Ectodomain
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX
;      STREET:  1100 New York Ave., NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,199
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,862
;      FILING DATE:  13-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cimbala, Michele A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1102.0130001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
```

```
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4378..4443
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  22026..22107
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  23002..23483
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  23905..24040
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  24252..24418
US-08-488-199-5

  Query Match             26.7%;  Score 16;  DB 2;  Length 26700;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 CTCCGGACCAGGACAT 44
               ||||||||||||||||
Db       3800 CTCCGGACCAGGACAT 3815


RESULT 10
US-08-760-534A-1
; Sequence 1, Application US/08760534A
; Patent No. 6017727
;  GENERAL INFORMATION:
;    APPLICANT:  JALKANEN, MARKKU
;    APPLICANT:  JAAKKOLA, PANU
;    APPLICANT:  VIHINEN, TAPANI
;    TITLE OF INVENTION:  SYNDECAN ENHANCER ELEMENT AND SYNDECAN
;    TITLE OF INVENTION:  STIMULATION OF CELLULAR DIFFERENTIATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/760,534A
;      FILING DATE:  02-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,186
;      FILING DATE:  07-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI93/00514
;      FILING DATE:  01-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIMBALA, MICHELE A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1708.0050004/MAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4378..4443, 22026..22106, 23001..23483,
;      LOCATION:  23905..24039, 24251..24418)
US-08-760-534A-1

  Query Match             26.7%;  Score 16;  DB 3;  Length 26700;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 CTCCGGACCAGGACAT 44
               ||||||||||||||||
Db       3800 CTCCGGACCAGGACAT 3815


RESULT 11
US-09-336-757-1
; Sequence 1, Application US/09336757
; Patent No. 6492344
;  GENERAL INFORMATION:
;    APPLICANT:  JALKANEN, MARKKU
;    APPLICANT:  JAAKKOLA, PANU
;    APPLICANT:  VIHINEN, TAPANI
;    TITLE OF INVENTION:  SYNDECAN ENHANCER ELEMENT AND SYNDECAN
;    TITLE OF INVENTION:  STIMULATION OF CELLULAR DIFFERENTIATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/336,757
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/760,534
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI93/00514
;      FILING DATE:  01-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIMBALA, MICHELE A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1708.0050004/MAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4378..4443, 22026..22106, 23001..23483,
;      LOCATION:  23905..24039, 24251..24418)
US-09-336-757-1
```

```
  Query Match             26.7%;  Score 16;  DB 4;  Length 26700;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         29 CTCCGGACCAGGACAT 44
              ||||||||||||||||
Db       3800 CTCCGGACCAGGACAT 3815


RESULT 12
US-09-270-767-3339
; Sequence 3339, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3339
;   LENGTH: 540
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
;   OTHER INFORMATION: n means any nucleotide
US-09-270-767-3339

  Query Match             25.0%;  Score 15;  DB 4;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         10 GCAGCTATGGAACTG 24
              |||||||||||||||
Db        134 GCAGCTATGGAACTG 148


RESULT 13
US-09-270-767-18621
; Sequence 18621, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18621
;   LENGTH: 540
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
;   OTHER INFORMATION: n means any nucleotide
US-09-270-767-18621

  Query Match             25.0%;  Score 15;  DB 4;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         10 GCAGCTATGGAACTG 24
              |||||||||||||||
Db        134 GCAGCTATGGAACTG 148


RESULT 14
US-08-978-589A-1
; Sequence 1, Application US/08978589A
; Patent No. 6087145
;  GENERAL INFORMATION:
;    APPLICANT:  ISHII, Takeshi
;    APPLICANT:  MITSUDA, Satoshi
;    TITLE OF INVENTION:  ESTERASE GENE AND ITS USE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH, LLP
;      STREET:  P.O. BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VIRGINIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  22040
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,589A
;      FILING DATE:  26-NOV-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  20-4336P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1089 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Burckholderia cepacia
;      STRAIN:  SC-20
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1089
US-08-978-589A-1

  Query Match             25.0%;  Score 15;  DB 3;  Length 1089;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         46 CGTGAAGCAGGTGCT 60
              |||||||||||||||
Db        336 CGTGAAGCAGGTGCT 350


RESULT 15
US-09-336-601-2
; Sequence 2, Application US/09336601
; Patent No. 6184008
; GENERAL INFORMATION:
;  APPLICANT: OHTA, Hiromichi
;  APPLICANT: SUGAI, Takeshi
;  APPLICANT: ISHII, Takeshi
;  APPLICANT: MITSUDA, Satoshi
;  TITLE OF INVENTION: PRODUCTION OF OPTICALLY ACTIVE SPHINGOID COMPOUND
;  FILE REFERENCE: 2185-349P
;  CURRENT APPLICATION NUMBER: US/09/336,601
;  CURRENT FILING DATE: 1999-06-21
;  EARLIER APPLICATION NUMBER: 09/034,007
;  EARLIER FILING DATE: 1998-03-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1089
;   TYPE: DNA
;   ORGANISM: E. coli JM 109/pAL 612 strain
US-09-336-601-2
```

Query Match 25.0%; Score 15; DB 3; Length 1089;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          46 CGTGAAGCAGGTGCT 60
               |||||||||||||||
Db         336 CGTGAAGCAGGTGCT 350
```

Search completed: December 23, 2004, 01:33:36
Job time : 50.3258 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 23:36:53 ; Search time 841.086 Seconds
(without alignments)
397.214 Million cell updates/sec

Title: US-09-898-616A-2
Perfect score: 60
Sequence: 1 agctacgaagcagctatgga.....acatgcgtgaagcaggtgct 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4105333 seqs, 2784095677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8210666

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	60	100.0	60	9	US-09-861-688-7	Sequence 7, Appli
	2	60	100.0	60	10	US-09-898-616A-2	Sequence 2, Appli
	3	60	100.0	60	15	US-10-187-498A-2	Sequence 2, Appli
	4	60	100.0	60	16	US-10-647-371-6	Sequence 6, Appli
c	5	37	61.7	60	9	US-09-861-688-12	Sequence 12, Appl
c	6	37	61.7	60	10	US-09-898-616A-7	Sequence 7, Appli
c	7	37	61.7	60	10	US-09-898-616A-8	Sequence 8, Appli
c	8	37	61.7	60	15	US-10-187-498A-7	Sequence 7, Appli
c	9	37	61.7	60	15	US-10-187-498A-8	Sequence 8, Appli
c	10	37	61.7	60	16	US-10-647-371-11	Sequence 11, Appl
c	11	18	30.0	59	9	US-09-861-688-11	Sequence 11, Appl
c	12	18	30.0	59	10	US-09-898-616A-6	Sequence 6, Appli
c	13	18	30.0	59	15	US-10-187-498A-6	Sequence 6, Appli
c	14	18	30.0	59	16	US-10-647-371-10	Sequence 10, Appl
	15	18	30.0	432	9	US-09-920-300A-1284	Sequence 1284, Ap
	16	18	30.0	432	13	US-10-033-528-1284	Sequence 1284, Ap
	17	18	30.0	432	15	US-10-099-926-1284	Sequence 1284, Ap
	18	18	30.0	489	10	US-09-918-995-23144	Sequence 23144, A
	19	18	30.0	1717	15	US-10-171-581-9	Sequence 9, Appli
	20	18	30.0	3252	18	US-10-723-860-5226	Sequence 5226, Ap
	21	18	30.0	3299	15	US-10-006-285-513	Sequence 513, App
c	22	17	28.3	467	13	US-10-027-632-59298	Sequence 59298, A
c	23	17	28.3	467	13	US-10-027-632-59299	Sequence 59299, A
c	24	17	28.3	467	13	US-10-027-632-59300	Sequence 59300, A
c	25	17	28.3	467	13	US-10-027-632-298628	Sequence 298628,
c	26	17	28.3	467	13	US-10-027-632-298629	Sequence 298629,
c	27	17	28.3	467	13	US-10-027-632-298630	Sequence 298630,
c	28	17	28.3	467	15	US-10-027-632-59298	Sequence 59298, A
c	29	17	28.3	467	15	US-10-027-632-59299	Sequence 59299, A
c	30	17	28.3	467	15	US-10-027-632-59300	Sequence 59300, A
c	31	17	28.3	467	15	US-10-027-632-298628	Sequence 298628,
c	32	17	28.3	467	15	US-10-027-632-298629	Sequence 298629,
c	33	17	28.3	467	15	US-10-027-632-298630	Sequence 298630,
	34	17	28.3	534	13	US-10-027-632-50093	Sequence 50093, A
	35	17	28.3	534	15	US-10-027-632-50093	Sequence 50093, A
c	36	17	28.3	555	13	US-10-027-632-321780	Sequence 321780,
c	37	17	28.3	555	15	US-10-027-632-321780	Sequence 321780,
	38	16	26.7	65	10	US-09-908-975-28883	Sequence 28883, A
	39	16	26.7	351	17	US-10-430-201-2182	Sequence 2182, Ap
	40	16	26.7	351	17	US-10-430-201-2183	Sequence 2183, Ap
	41	16	26.7	373	9	US-09-983-965-739	Sequence 739, App
	42	16	26.7	375	18	US-10-674-124A-22078	Sequence 22078, A
c	43	16	26.7	477	9	US-09-783-590-4830	Sequence 4830, Ap
	44	16	26.7	492	18	US-10-425-115-102427	Sequence 102427,
	45	16	26.7	561	13	US-10-027-632-10327	Sequence 10327, A

ALIGNMENTS

RESULT 1
US-09-861-688-7
; Sequence 7, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-7

Query Match 100.0%; Score 60; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 6.3e-24;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60

RESULT 2
US-09-898-616A-2

```
; Sequence 2, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-2

  Query Match             100.0%;  Score 60;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-24;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60


RESULT 3
US-10-187-498A-2
; Sequence 2, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-2

  Query Match             100.0%;  Score 60;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-24;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60


RESULT 4
US-10-647-371-6
; Sequence 6, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-6

  Query Match             100.0%;  Score 60;  DB 16;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-24;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACCAGGACATGCGTGAAGCAGGTGCT 60


RESULT 5
US-09-861-688-12/c
; Sequence 12, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
;  TITLE OF INVENTION:  Inflammatory and
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE: 116142/2
;  CURRENT APPLICATION NUMBER: US/09/861,688
;  CURRENT FILING DATE:  2001-05-21
;  PRIOR APPLICATION NUMBER: 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-09-861-688-12

  Query Match             61.7%;  Score 37;  DB 9;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 6
US-09-898-616A-7/c
; Sequence 7, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-7

  Query Match             61.7%;  Score 37;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 7
US-09-898-616A-8/c
; Sequence 8, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-8

  Query Match             61.7%;  Score 37;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 8
US-10-187-498A-7/c
; Sequence 7, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-7

  Query Match             61.7%;  Score 37;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 9
US-10-187-498A-8/c
; Sequence 8, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-8

  Query Match             61.7%;  Score 37;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 10
US-10-647-371-11/c
; Sequence 11, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-11

  Query Match           61.7%;  Score 37;  DB 16;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 AGCTACGAAGCAGCTATGGAACTGTTCTCTCCGGACC 1


RESULT 11
US-09-861-688-11/c
; Sequence 11, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
;  TITLE OF INVENTION:  Inflammatory and
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE: 116142/2
;  CURRENT APPLICATION NUMBER: US/09/861,688
;  CURRENT FILING DATE:  2001-05-21
;  PRIOR APPLICATION NUMBER: 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-09-861-688-11

  Query Match           30.0%;  Score 18;  DB 9;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42


RESULT 12
US-09-898-616A-6/c
; Sequence 6, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-6

  Query Match           30.0%;  Score 18;  DB 10;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42


RESULT 13
US-10-187-498A-6/c
; Sequence 6, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-6

  Query Match           30.0%;  Score 18;  DB 15;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42
```

```
RESULT 14
US-10-647-371-10/c
; Sequence 10, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-10

  Query Match            30.0%;  Score 18;  DB 16;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 3.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         38 AGGACATGCGTGAAGCAG 55
              ||||||||||||||||||
Db         59 AGGACATGCGTGAAGCAG 42


RESULT 15
US-09-920-300A-1284
; Sequence 1284, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
;  APPLICANT: King, Gordon E.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.547
;  CURRENT APPLICATION NUMBER: US/09/920,300A
;  CURRENT FILING DATE:  2001-07-31
;  NUMBER OF SEQ ID NOS: 1789
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1284
;   LENGTH: 432
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-920-300A-1284

  Query Match            30.0%;  Score 18;  DB 9;  Length 432;
  Best Local Similarity  100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         41 ACATGCGTGAAGCAGGTG 58
              ||||||||||||||||||
Db          9 ACATGCGTGAAGCAGGTG 26


Search completed: December 23, 2004, 05:19:26
Job time : 843.086 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:17:58 ; Search time 834.543 Seconds
(without alignments)
3343.258 Million cell updates/sec

Title: US-09-898-616A-3
Perfect score: 59
Sequence: 1 cagctgaagaaactggttga.....cgtgaatccatcataaactg 59

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		59	100.0	59	6	AR160917	AR160917 Sequence
c 2		37	62.7	59	6	AR160920	AR160920 Sequence
3		19	32.2	62318	2	AC135267	AC135267 Rattus no
4		19	32.2	65394	2	AC117645	AC117645 Mus muscu
5		19	32.2	145825	8	OSJN00027	AL606600 Oryza sat
6		19	32.2	232344	2	AC111782	AC111782 Rattus no
c 7		19	32.2	246787	2	AC128608	AC128608 Rattus no
8		18	30.5	215	6	A13577	A13577 IGF-II gene
c 9		18	30.5	215	6	A13578	A13578 IGF-II gene
10		18	30.5	220	12	SYNBILGFG	M60420 Artificial
c 11		18	30.5	1234	8	AK073198	AK073198 Oryza sat
c 12		18	30.5	2004	6	AX461306	AX461306 Sequence
13		18	30.5	3176	8	AK065797	AK065797 Oryza sat
14		18	30.5	7419	1	PSETRPEGDC	M33799 P.putida tr
c 15		18	30.5	42499	8	AC004625	AC004625 Arabidops
16		18	30.5	67388	9	AC133476	AC133476 Homo sapi
17		18	30.5	96066	8	AC005662	AC005662 Arabidops
c 18		18	30.5	106057	8	AC119796	AC119796 Oryza sat
19		18	30.5	110000	1	RME591985_15	Continuation (16 o
c 20		18	30.5	150503	8	AC122149	AC122149 Oryza sat
c 21		18	30.5	159867	2	AC130411	AC130411 Homo sapi
c 22		18	30.5	166399	2	AC012584	AC012584 Homo sapi
23		18	30.5	169199	2	AC109292	AC109292 Mus muscu
c 24		18	30.5	170064	8	AP003771	AP003771 Oryza sat
c 25		18	30.5	193386	10	AC117693	AC117693 Mus muscu
c 26		18	30.5	202920	10	AL929441	AL929441 Mouse DNA
27		18	30.5	270725	2	AC126948	AC126948 Rattus no
c 28		17	28.8	195	6	AX435021	AX435021 Sequence
c 29		17	28.8	993	13	AY487489	AY487489 Unculture
30		17	28.8	2190	10	AF144255	AF144255 Mus muscu
31		17	28.8	3530	6	CQ582203	CQ582203 Sequence
c 32		17	28.8	5568	6	AX571642	AX571642 Sequence
c 33		17	28.8	10166	1	AE007324	AE007324 Streptoco
c 34		17	28.8	11627	1	AE012368	AE012368 Xanthomon
c 35		17	28.8	14736	6	CQ789080	CQ789080 Sequence
c 36		17	28.8	14736	6	AR218939	AR218939 Sequence
c 37		17	28.8	14736	6	BD003851	BD003851 Polynucle
c 38		17	28.8	19457	2	AC017722	AC017722 Drosophil
c 39		17	28.8	62598	6	AX571767	AX571767 Sequence
40		17	28.8	66904	2	AC104954	AC104954 Homo sapi
c 41		17	28.8	91803	9	AC092318	AC092318 Homo sapi
c 42		17	28.8	95506	9	AC002407	AC002407 Human Chr
c 43		17	28.8	100061	2	AC016551	AC016551 Homo sapi
44		17	28.8	132164	9	AC098587	AC098587 Homo sapi
c 45		17	28.8	144402	2	AC021854	AC021854 Homo sapi

ALIGNMENTS

RESULT 1
AR160917
LOCUS AR160917 59 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 8 from patent US 6255281.
ACCESSION AR160917
VERSION AR160917.1 GI:16225984
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 59)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
TITLE Use of recombinant human uteroglobin in treatment of inflammatory and fibrotic conditions
JOURNAL Patent: US 6255281-A 8 03-JUL-2001;
FEATURES Location/Qualifiers
source 1. .59
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 59; DB 6; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.3e-23;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59

RESULT 2
AR160920/c
LOCUS AR160920 59 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 11 from patent US 6255281.
ACCESSION AR160920
VERSION AR160920.1 GI:16225993
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 59)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.

```
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 11 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .59
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             62.7%;  Score 37;  DB 6;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-10;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 3
AC135267
LOCUS       AC135267               62318 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-416G11, WORKING DRAFT SEQUENCE.
ACCESSION   AC135267
VERSION     AC135267.2  GI:25073198
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 62318)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwackelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62318)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 62318)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23802942.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
               Center project name: KCYH
               Center clone name: CH230-416G11
            ---------- Summary Statistics
               Assembly program: Phrap; version 0.990329
               Consensus quality: 52773 bases at least Q40
               Consensus quality: 53668 bases at least Q30
               Consensus quality: 54142 bases at least Q20
               Estimated insert size: 52859; sum-of-contigs estimation
               Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1    62318: contig of 62318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .62318
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-416G11"
     misc_feature    1. .1459
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(7376. .8361)
                     /note="clone_boundary
```

```
                     clone_end:Sp6
                     site:
                     end_sequence:BZ195619"
     misc_feature    complement(54571. .55515)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ195617"
     misc_feature    60931. .62318
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             32.2%;  Score 19;  DB 2;  Length 62318;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CTGAAGAAACTGGTTGACA 22
              ||||||||||||||||||
Db       9891 CTGAAGAAACTGGTTGACA 9909


RESULT 4
AC117645
LOCUS       AC117645               65394 bp    DNA     linear   HTG 30-AUG-2002
DEFINITION  Mus musculus clone RP23-238F4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC117645
VERSION     AC117645.5  GI:22549859
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 65394)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-238F4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65394)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 65394)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:22474966.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23587
                Center clone name: 238_F_4
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      727: contig of 727 bp in length
            *      728      827: gap of 100 bp
            *      828     1548: contig of 721 bp in length
            *     1549     1648: gap of 100 bp
            *     1649     2369: contig of 721 bp in length
            *     2370     2469: gap of 100 bp
            *     2470     3184: contig of 715 bp in length
            *     3185     3284: gap of 100 bp
            *     3285     3977: contig of 693 bp in length
            *     3978     4077: gap of 100 bp
            *     4078     4796: contig of 719 bp in length
            *     4797     4896: gap of 100 bp
            *     4897     5612: contig of 716 bp in length
            *     5613     5712: gap of 100 bp
            *     5713     6431: contig of 719 bp in length
            *     6432     6531: gap of 100 bp
            *     6532     7243: contig of 712 bp in length
            *     7244     7343: gap of 100 bp
            *     7344     8046: contig of 703 bp in length
            *     8047     8146: gap of 100 bp
            *     8147     8866: contig of 720 bp in length
            *     8867     8966: gap of 100 bp
            *     8967     9684: contig of 718 bp in length
            *     9685     9784: gap of 100 bp
            *     9785    10492: contig of 708 bp in length
            *    10493    10592: gap of 100 bp
            *    10593    11318: contig of 726 bp in length
            *    11319    11418: gap of 100 bp
            *    11419    12131: contig of 713 bp in length
            *    12132    12231: gap of 100 bp
            *    12232    12938: contig of 707 bp in length
            *    12939    13038: gap of 100 bp
            *    13039    13748: contig of 710 bp in length
            *    13749    13848: gap of 100 bp
            *    13849    14557: contig of 709 bp in length
            *    14558    14657: gap of 100 bp
            *    14658    15372: contig of 715 bp in length
```

```
*    15373    15472: gap of 100 bp
*    15473    16176: contig of 704 bp in length
*    16177    16276: gap of 100 bp
*    16277    16968: contig of 692 bp in length
*    16969    17068: gap of 100 bp
*    17069    17778: contig of 710 bp in length
*    17779    17878: gap of 100 bp
*    17879    18599: contig of 721 bp in length
*    18600    18699: gap of 100 bp
*    18700    19406: contig of 707 bp in length
*    19407    19506: gap of 100 bp
*    19507    20190: contig of 684 bp in length
*    20191    20290: gap of 100 bp
*    20291    20999: contig of 709 bp in length
*    21000    21099: gap of 100 bp
*    21100    21804: contig of 705 bp in length
*    21805    21904: gap of 100 bp
*    21905    22614: contig of 710 bp in length
*    22615    22714: gap of 100 bp
*    22715    23406: contig of 692 bp in length
*    23407    23506: gap of 100 bp
*    23507    24226: contig of 720 bp in length
*    24227    24326: gap of 100 bp
*    24327    25041: contig of 715 bp in length
*    25042    25141: gap of 100 bp
*    25142    25859: contig of 718 bp in length
*    25860    25959: gap of 100 bp
*    25960    26665: contig of 706 bp in length
*    26666    26765: gap of 100 bp
*    26766    27478: contig of 713 bp in length
*    27479    27578: gap of 100 bp
*    27579    28296: contig of 718 bp in length
*    28297    28396: gap of 100 bp
*    28397    29079: contig of 683 bp in length
*    29080    29179: gap of 100 bp
*    29180    29888: contig of 709 bp in length
*    29889    29988: gap of 100 bp
*    29989    30692: contig of 704 bp in length
*    30693    30792: gap of 100 bp
*    30793    31504: contig of 712 bp in length
*    31505    31604: gap of 100 bp
*    31605    32316: contig of 712 bp in length
*    32317    32416: gap of 100 bp
*    32417    33128: contig of 712 bp in length
*    33129    33228: gap of 100 bp
*    33229    33944: contig of 716 bp in length
*    33945    34044: gap of 100 bp
*    34045    34757: contig of 713 bp in length
*    34758    34857: gap of 100 bp
*    34858    35572: contig of 715 bp in length
*    35573    35672: gap of 100 bp
*    35673    36386: contig of 714 bp in length
*    36387    36486: gap of 100 bp
*    36487    37196: contig of 710 bp in length
*    37197    37296: gap of 100 bp
*    37297    37993: contig of 697 bp in length
*    37994    38093: gap of 100 bp
*    38094    38805: contig of 712 bp in length
*    38806    38905: gap of 100 bp
*    38906    39615: contig of 710 bp in length
*    39616    39715: gap of 100 bp
*    39716    40424: contig of 709 bp in length
*    40425    40524: gap of 100 bp
*    40525    41216: contig of 692 bp in length
*    41217    41316: gap of 100 bp
*    41317    42027: contig of 711 bp in length
*    42028    42127: gap of 100 bp
*    42128    42837: contig of 710 bp in length
*    42838    42937: gap of 100 bp
*    42938    43653: contig of 716 bp in length
*    43654    43753: gap of 100 bp
*    43754    44461: contig of 708 bp in length
*    44462    44561: gap of 100 bp
*    44562    45267: contig of 706 bp in length
*    45268    45367: gap of 100 bp
*    45368    46054: contig of 687 bp in length
*    46055    46154: gap of 100 bp
*    46155    46866: contig of 712 bp in length

  Query Match           32.2%;  Score 19;  DB 2;  Length 65394;
  Best Local Similarity  100.0%;  Pred. No. 5.9;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         17 TTGACACCCTGCCGCAGAA 35
              |||||||||||||||||||
Db      42513 TTGACACCCTGCCGCAGAA 42531


RESULT 5
OSJN00027
LOCUS       OSJN00027              145825 bp    DNA     linear   PLN 10-FEB-2004
DEFINITION  Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0067G20,
            complete sequence.
ACCESSION   AL606600
VERSION     AL606600.3  GI:32479982
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,
            Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,
            Weng,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,
            Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,
            Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,
            Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,
            Zhang,Y., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,
            Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,
            Ni,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,
            Li,J., Hong,G., Xue,Y. and Han,B.
  TITLE     Sequence and analysis of rice chromosome 4
  JOURNAL   Nature 420 (6913), 316-320 (2002)
  MEDLINE   22337377
   PUBMED   12447439
REFERENCE   2
  AUTHORS   Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,
            Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,
            Chen,X.Y., Shao,Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W.,
            Wang,L.J., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,
            Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,
            Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,
            Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y.,
            Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,
            Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.
            and Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2001) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBa0067G20.
COMMENT     On Jul 8, 2003 this sequence version replaced gi:21740578.
            Web site: http://www.ncgr.ac.cn
            -------------- Summary Statistics
            Assembly program: phrap
            --------------
            This is a complete sequence.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM
            (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean
            Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
            complete sequence against NCBI none redundant protein database (nr)
```

(ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.

```
FEATURES             Location/Qualifiers
     source          1. .145825
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="4"
                     /clone="OSJNBa0067G20"
                     /clone_lib="CUGI-OSJNBa"
     gene            40. .2473
                     /gene="OSJNBa0067G20.1"
     CDS             join(40. .679,759. .883,951. .1097,1204. .1262,1330. .1575,
                     1766. .1884,1983. .2473)
                     /gene="OSJNBa0067G20.1"
                     /codon_start=1
                     /protein_id="CAD40485.2"
                     /db_xref="GI:38345106"
                     /db_xref="TrEMBL:Q7X7F5"
                     /translation="MEREAGGTNPQPVLGRMVVIEDYILCDFIPPPSEFLLLVLNFYG
                     LSLLHLNPNSIAFLSIFSHLCEAYIGVEPFLDLFRFYYELRWMEPNRVSGCVGFRLRD
                     GLKSRYIPFQCPSSRSKWRTRWFYLQIKDSDPVFIVPTEQPDKIPSWTAKPPLTPSLQ
                     SFIDIVDDLRVRGLSRYEVTADFVRRRIQPLQARVHPAFDYSGPEDATRVSPQGLNSE
                     TVGRRVGQVMISSPTTASNIPVPLCEKRAAERDAAINALPLTNIIGPLADHQVAASLK
                     EKVAKEASDAAAATTSGGNVLRKGRKFSASDASPLPPRRQRLVTVGEKAARAKAVQDE
                     SGGTSGASPAAASTDVVLAPGSREVTPSGPVSDPAAGRGPLAAVLTWEDLQVEMGRLL
                     EASARGIGRDGRSRDVDDEDVQQRGQPEFEPHASPDNVDQTLATGPEMVTTSLDLESA
                     LGGQDRRIQYWKTKFEVADLERTMLAVKKEQAVETLRGGEVWFNSYLKSCCTSMARVC
                     RELRVPHGDPEESAAGYISWLNGACAQLDGVGKRIDEALKQECHRSSRYAGGHVLACL
                     RDLREGFARSRRTPAEIDHLARSMAPFAEKIFQSMDWRWPSW"
     gene            complement(3453. .6561)
                     /gene="OSJNBa0067G20.2"
     CDS             complement(join(3453. .4786,4844. .5969,6199. .6561))
                     /gene="OSJNBa0067G20.2"
                     /codon_start=1
                     /protein_id="CAD40484.2"
                     /db_xref="GI:38345107"
                     /db_xref="TrEMBL:Q7X8E5"
                     /translation="MAASHYAYQVLKMSGPKGTITIQGNTKLAVQCDKRSLDMVEQTP
                     SPPDTTEPPKKVSKTNKTLKPDGAIKIVPLYSANHDKTVKIGAILSEKLELALITFLR
                     DNADVFAWQPSDMPGVPREACPKDHFPLARIDQLVDSTAGCELLSFLDAYSGYHQISM
                     AKEDEEKTTFITPFGVFCYVKMPFELITAGNTFQRTVQGALSDQLGNNVEAYVDDIVV
                     KIKTCDSLIDDLRETFDNLRRYHLMLNREKCTFEVPSGKLLGFLVSGRGIEANPEKIK
                     AIENMKSPTRLKEVQKLIGCMAALSRFVARMGERGQPPLALLKKQDKFVWTQEAEEAF
                     VALKRYLSNPPVLVAPQPNEELFLYIAATPYSVSTIIVVEREKVQRPVYYVSEALHDA
                     KTRYPQIQKLLYAVIMTSRKLRHYFQAHRVTVVSSFPLGEVVRNKDVVGRIAKWIVEL
                     SQFDVHFAPRTAIKSQVLADFVADWTMPDNKSDNQLDNETWTMAFDDQFKHVIHLNFR
                     VTNNTAEYEGLLARIRAAATLGVKRLIVKGASELVANQVHKDYKCSNPELSKYLVEVR
                     KLEKRFDGIEVRHVYRKDNIELDDLARRASRQEPLEPGTFLDILTKPSVKEVSGEVSP
                     VTLDISSKATKAERAVADIETTDDWRFPLIKFINSEELPEDDTEAEKITRKAKIYCMV
                     GNNLYKKAPNRVLLKSISSDNGRHLLLDIHEGIYGSHAAGWTLVGKAFRQGFFWPTTL
                     KYACDMVQRCEACQFHSKHTRLSAQALQTIPLTWPFSCWGLDILGPFPRGQGGYRFQF
                     VAIDKFTKWIEAVPMGEIKAVNAIKFIKGIFCRYGLPHRIITDNGSQFISADFQDYCI
                     GLGVKICFASVSHPQSNGQVERANGIVLQGIKTRVYDRLMSHDKKWVEELPSVLWAVR
                     TTPTTCNKETTFFLVYGSEAMLPSEL"
     gene            6833. .9625
                     /gene="OSJNBa0067G20.3"
     CDS             join(6833. .7754,7830. .8701,8774. .9625)
                     /gene="OSJNBa0067G20.3"
                     /codon_start=1
                     /protein_id="CAD40483.1"
                     /db_xref="GI:21740600"
                     /db_xref="TrEMBL:Q7X665"
                     /translation="MAANELAEVKKQLEELKEKGYIRPSTSPQGAPVIFVEKKDKTKR
                     MCVDYRALNEVTIKNKYPLPRIDDLFDQLKGATVFSKIDLRSGYHQLRIREEDIPKTA
                     FTTRYGLYEFTVMSFGLTNAPAFPMNLMNKVFMEYLDKFVVVFIDDILVYSQSEEDHQ
                     QHLRLVLGKLREHQLYAKLRKCEFCLSEVKFLGHVISAKGVAVDPKTVTAVTDWKQPK
                     TVTQVRSFLGLAGYYRRFIENFSKIAQPMTQLLKKEEKFVWSPQCEKAFQTLKEKLVS
                     SPVLILPDTRKDFLVYCDASRQGLGRVLMQEAAVVHALKIWRHYLIGNRCEIYTDHKS
                     LKYIFTQSDLNLRQRRWLELIKDYDVRIHYHPGKANVVADPLSRKSHCNTLGVRGIPP
                     ELNQQMEALNLSIVGRGFLATLEAKPTLLDQIREAQKNDPDMYGLLKNMKQGKAAGFT
                     EDEHGTLWNGNRVCVPDDWELKQLILQEAHESSYSIHPGSTKIYLDLKEKYWWISMKR
                     EIAEFVALCDVCQHVKAEHQRPARLLQPLQVPEWKWDEIGMDFITGLPKTQGGYDSIW
                     VVVDRLTKVARFIPVKTTYGGNKLAELYFARIKLQEELGTRLNFSTAYHPQTDGQTER
                     LNQILEDILRACVLNFGKTWDKSLPYAEFSYNNSYQASIQMAPYEALYGRKCQTPLLW
                     DQVGESQVFGTDILREAEVKVRTIQENLKVAQSRQKSYADNRRRNLEFAVDNFVYLRV
                     TPLRGVHRFQTKGKLAPRFVGPFRIIARRGEVAYQLELPASLGNVHDVFHLSQLKKCL
                     RVPSEQADSEQIEVREDLTYVERPVKILDTMERRTRNRVIRFYKVQWSNHAEEEATWE
                     REDELKAAHPDLFASSSESRGRDSV"
     gene            complement(13199. .15084)
                     /gene="OSJNBa0067G20.4"
     CDS             complement(join(13199. .13606,13767. .14071,14133. .15084))
                     /gene="OSJNBa0067G20.4"
                     /codon_start=1
                     /protein_id="CAD40482.1"
                     /db_xref="GI:21740599"
                     /db_xref="TrEMBL:Q7X7S6"
                     /translation="MSSSLVNNSVNTNLSQVFPASTMLVWTQVGEIVFPVYTTMPIST
                     CPLMTGNENAVATNRDDSMSKDTPTEAKNGISTTSELEKDSNAAKPCRSDMNREPTKM
                     TSEATRSWCPIHKTSKHTLQTCWVFLNVRAEIRICKERGIQRISPTRDVYCPIHKTKN
                     HDLSSCKVLLSAMKAPPPKVQQSHIPIRDKDKERGATPTSDRFVGVIDIDPHEPSVLH
                     LLEDYGSSTTSALREVLAIDDVGTSARTNAEAENQLATPAQHIRAVNAILRETPYDPV
                     LNDDLARWTARLRESVTNLSNTFEEAAAAAHPEQPPTGDANAISEITSTAAERHDAHE
                     TMKIVPGVMSLHGAIMTKIEETAQTKTEAVITATTATITTIANGGCQTILVEDVAITT
                     TMMETGAGTTAGDDDRILEIQADILVIRFRPGAIEKYDGSTDSEEFLQVYFTVLYAAG
                     ADDNALANYLPTALKGSARSWLMHLLPYSISSWADLWQQFVANFQGTYKRHAIEDDLH
                     ALTQNSGESLREYVRRFNECRNTIPEITDASVIHAFKSGVRDCYTT"
     gene            complement(19850. .20257)
                     /gene="OSJNBa0067G20.5"
     CDS             complement(19850. .20257)
                     /gene="OSJNBa0067G20.5"
                     /codon_start=1
                     /protein_id="CAD40481.2"
                     /db_xref="GI:38345108"
                     /db_xref="TrEMBL:Q7X8D8"
                     /translation="MAALLQLAMSVAAVPHLPEEETTPSVDVEIASPDQQHLAAAAAP
                     SMAVFPPAADEEEEEEAVAVLLSGEFFTCWSSLSESTDDETTRRTTTESMFYIPTNGS
                     NDVGRRRRKVRSWSRGSFLGRGSFGMVFEGITK"
     gene            complement(21613. .22857)
                     /gene="OSJNBa0067G20.6"
     CDS             complement(join(21613. .22296,22777. .22857))
                     /gene="OSJNBa0067G20.6"
                     /codon_start=1
                     /protein_id="CAD40480.1"
                     /db_xref="GI:21740597"
                     /db_xref="TrEMBL:Q7XVR8"
                     /translation="MQLNPKARTRRNKTKAKGWQCAEIVLNVLADLFSFDIKDYFDET
                     KEDTTSKALAPLSDDVKRTLDDISHRMEASSLDSLVVDCGPIRTRLHEVQALIPDELA
                     DVLTPAVYLKQQQFKLEKAKLRLAERRERKDIEATIQANRQLVHEEKAKLDQLSEGPI
                     KSNIDRLEARKIELLAQLQECNTELDMEHKKLADLPQAVEEQKARLKSAIKNVADMTK
                     SLKVIPGTDAQDAQAIEEVEQIRQRAISAIQRYLSQ"
     gene            23570. .24858
                     /gene="OSJNBa0067G20.7"
     CDS             join(23570. .23705,23930. .24120,24214. .24858)
                     /gene="OSJNBa0067G20.7"
                     /codon_start=1
                     /protein_id="CAE01610.1"
                     /db_xref="GI:32479983"
                     /db_xref="TrEMBL:Q7XT96"
                     /translation="MREQFGLKPKETRSLYRQPYPEWFERVPLPNRFKVLDFSKFSGQ
                     DAIKQRYDEPVHEYIQRPREMRNKCFSLSLTDAQLADLAFQGMIAPIREKFSSEDFDS
                     LSHLIQKKDDSEIAAAEWVRSKKVISCQWVKSPRKEKKYDFNITKADKIFDLLLREKQ
                     IQLPAGHIIPSAEELGKKRYCKWHNSGSHSTNDCKVFRQQIQVAIEGGKIRFDDSKKP
                     MKVDDNPFPVNMVHTAGQIADRVRAKGFQVNSAKIINKYQRKYDKQQEKHYEEYDDGF
                     DPHWGCEFFRLCWNEGMRLCHAQKFTSNFRTYLCIKSSSRNQPRYTN"
     gene            complement(32437. .41228)
                     /gene="OSJNBa0067G20.8"
     CDS             complement(join(32437. .32730,33273. .33521,35840. .35975,
                     38583. .38801,40070. .40448,40988. .41228))
                     /gene="OSJNBa0067G20.8"
                     /codon_start=1
                     /protein_id="CAE01611.2"
                     /db_xref="GI:38345109"

Query Match          32.2%;  Score 19;  DB 8;  Length 145825;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 AGCTGAAGAAACTGGTTGA 20
              ||||||||||||||||||||
Db     101851 AGCTGAAGAAACTGGTTGA 101869


RESULT 6
AC111782
LOCUS       AC111782               232344 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-226J1, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC111782
VERSION     AC111782.4  GI:25073626
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 232344)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232344)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 232344)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23603299.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOCG
                Center clone name: CH230-226J1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222543 bases at least Q40
                Consensus quality: 224732 bases at least Q30
                Consensus quality: 226208 bases at least Q20
                Estimated insert size: 225765; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   226162: contig of 226162 bp in length
            *  226163   226262: gap of unknown length
            *  226263   230757: contig of 4495 bp in length
            *  230758   230857: gap of unknown length
            *  230858   232344: contig of 1487 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232344
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-226J1"
     misc_feature    1. .1538
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1589. .2718
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    4907. .5729
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ103653"
```

```
     misc_feature    222851. .224016
                     /note="wgs_contig"
     misc_feature    224067. .226162
                     /note="wgs_contig"
     misc_feature    226263. .227570
                     /note="wgs_contig"
     misc_feature    227922. .230757
                     /note="wgs_contig"
ORIGIN

  Query Match          32.2%;  Score 19;  DB 2;  Length 232344;
  Best Local Similarity   100.0%;  Pred. No. 5.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CTGAAGAAACTGGTTGACA 22
              ||||||||||||||||||
Db     191901 CTGAAGAAACTGGTTGACA 191919


RESULT 7
AC128608/c
LOCUS       AC128608               246787 bp    DNA     linear   HTG 09-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-320L21, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC128608
VERSION     AC128608.2  GI:22856092
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 246787)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 246787)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 246787)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21909412.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYJN
                Center clone name: CH230-320L21
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 219698 bases at least Q40
                Consensus quality: 223003 bases at least Q30
                Consensus quality: 224973 bases at least Q20
                Estimated insert size: 239901; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   245086: contig of 245086 bp in length
            *    245087   245186: gap of unknown length
            *    245187   246787: contig of 1601 bp in length.
FEATURES             Location/Qualifiers
     source          1. .246787
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-320L21"
     misc_feature    1. .1472
                     /note="wgs_contig"
     misc_feature    54443. .56179
                     /note="wgs_contig"
     misc_feature    241917. .243271
                     /note="wgs_contig"
     misc_feature    243322. .245086
```

```
                     /note="wgs_contig"
ORIGIN

  Query Match             32.2%;  Score 19;  DB 2;  Length 246787;
  Best Local Similarity   100.0%;  Pred. No. 5.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CTGAAGAAACTGGTTGACA 22
              |||||||||||||||||||
Db     210071 CTGAAGAAACTGGTTGACA 210053


RESULT 8
A13577
LOCUS       A13577                   215 bp    DNA     linear   PAT 31-DEC-1993
DEFINITION  IGF-II gene.
ACCESSION   A13577
VERSION     A13577.1  GI:491682
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Nygren,P.A., Abrahmsen,L. and Uhlen,M.
  TITLE     A recombinant fusion protein, its use and a recombinant vector
  JOURNAL   Patent: EP 0333691-A 1 20-SEP-1989;
            CEMU BIOTEKNIK AB
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     gene            7. .213
                     /gene="IGF II"
     CDS             7. .213
                     /gene="IGF II"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA01122.1"
                     /db_xref="GI:491683"
                     /translation="MAYRPSETLCGGELVDTLQFVCGDRGFYFSRPASRVSRRSRGIV
                     EECCFRSCDLALLETYCATPAKSE"
ORIGIN

  Query Match             30.5%;  Score 18;  DB 6;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 AACTGGTTGACACCCTGC 28
              ||||||||||||||||||
Db         44 AACTGGTTGACACCCTGC 61


RESULT 9
A13578/c
LOCUS       A13578                   215 bp    DNA     linear   PAT 31-DEC-1993
DEFINITION  IGF-II gene.
ACCESSION   A13578
VERSION     A13578.1  GI:489637
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Nygren,P.A., Abrahmsen,L. and Uhlen,M.
  TITLE     A recombinant fusion protein, its use and a recombinant vector
  JOURNAL   Patent: EP 0333691-A 2 20-SEP-1989;
            CEMU BIOTEKNIK AB
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             30.5%;  Score 18;  DB 6;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 AACTGGTTGACACCCTGC 28
              ||||||||||||||||||
Db        176 AACTGGTTGACACCCTGC 159


RESULT 10
SYNBILGFG
LOCUS       SYNBILGFG                220 bp    DNA     linear   SYN 27-APR-1993
DEFINITION  Artificial bovine insulin-like growth factor 2 gene, complete cds.
ACCESSION   M60420
VERSION     M60420.1  GI:208029
KEYWORDS    insulin-like growth factor II.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Easton,A.M., Gierse,J.K., Seetharam,R., Klein,B.K. and Kotts,C.E.
  TITLE     Production of bovine insulin-like growth factor 2 (bIGF2) in
            Escherichia coli
  JOURNAL   Gene 101 (2), 291-295 (1991)
  MEDLINE   91276286
   PUBMED   2055493
COMMENT     Original source text: Synthetic DNA.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
     CDS             7. .216
                     /note="bovine"
                     /codon_start=1
                     /transl_table=11
                     /product="insulin-like growth factor 2"
                     /protein_id="AAA72971.1"
                     /db_xref="GI:208030"
                     /translation="MAAYRPSETLCGGELVDTLQFVCGDRGFYFSRPSSRINRRSRGI
                     VEECCFRSCDLALLETYCATPAKSE"
ORIGIN

  Query Match             30.5%;  Score 18;  DB 12;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 AACTGGTTGACACCCTGC 28
              ||||||||||||||||||
Db         47 AACTGGTTGACACCCTGC 64


RESULT 11
AK073198/c
LOCUS       AK073198                 1234 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J033022I15, full
            insert sequence.
ACCESSION   AK073198
VERSION     AK073198.1  GI:32983221
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
```

Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 1234)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE Direct Submission

JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .1234
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033022I15"

ORIGIN

Query Match 30.5%; Score 18; DB 8; Length 1234;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         13 CTGGTTGACACCCTGCCG 30
              ||||||||||||||||||
Db        660 CTGGTTGACACCCTGCCG 643
```

RESULT 12
AX461306/c
LOCUS AX461306 2004 bp DNA linear PAT 08-JUL-2002
DEFINITION Sequence 235 from Patent WO0198480.
ACCESSION AX461306
VERSION AX461306.1 GI:21726514
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Budworth,P., Brown,D., Chang,H.S., Zhu,T., Han,B., Wang,X. and Cooper,B.
TITLE Promoters for regulation of plant gene expression
JOURNAL Patent: WO 0198480-A 235 27-DEC-2001;
Syngenta Participations AG (CH)
FEATURES Location/Qualifiers
source 1. .2004
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN

Query Match 30.5%; Score 18; DB 6; Length 2004;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         30 GCAGAAACCGCGTGAATC 47
              ||||||||||||||||||
Db        389 GCAGAAACCGCGTGAATC 372
```

RESULT 13
AK065797
LOCUS AK065797 3176 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013041K05, full insert sequence.
ACCESSION AK065797
VERSION AK065797.1 GI:32975815
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team:, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,

Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 3176)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE Direct Submission

JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers

source 1. .3176
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013041K05"

ORIGIN

Query Match 30.5%; Score 18; DB 8; Length 3176;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCTGAAGAAACTGGTTGA 20
||||||||||||||||||
Db 2681 GCTGAAGAAACTGGTTGA 2698

RESULT 14
PSETRPEGDC
LOCUS PSETRPEGDC 7419 bp DNA linear BCT 07-NOV-1995
DEFINITION P.putida trpEGDC operon encoding anthranilate synthetase alpha- and beta-subunits (trpE and trpG), phosphoribosyl transferase (trpD), and indoleglycerol phosphate synthetase (trpC) genes, complete cds.
ACCESSION M33799 M33800 M33801
VERSION M33799.1 GI:151627
KEYWORDS anthranilate synthetase; indoleglycerol phosphate synthetase; phosphoribosyl transferase.
SOURCE Pseudomonas putida
ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 7419)
AUTHORS Essar,D.W., Eberly,L. and Crawford,I.P.
TITLE Evolutionary differences in chromosomal locations of four early genes of the tryptophan pathway in fluorescent pseudomonads: DNA sequences and characterization of Pseudomonas putida trpE and trpGDC
JOURNAL J. Bacteriol. 172 (2), 867-883 (1990)
MEDLINE 90130325
PUBMED 2404959
COMMENT Original source text: P.putida (strain ATCC 232287) DNA.
FEATURES Location/Qualifiers

source 1. .7419
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="ATCC 232287"
/db_xref="taxon:303"
/clone="p14[14,45,50]"

gene 827. .2308
/gene="trpE"

CDS 827. .2308
/gene="trpE"
/codon_start=1
/transl_table=11
/product="anthranilate synthetase alpha-subunit"
/protein_id="AAA80551.1"
/db_xref="GI:151629"
/translation="MNREEFLRLAAVGYNRIPLACETLADFDTPLSIYLKLADQPNSY
LLESVQGGEKWGRYSMIGLPSRTVMRVHGYHVSILHDGVEVESHDVEDPLAFVESFKD
RYKVADIPGLPRFNGGLVGYFGYDCVRYVEKRLGVSPNPDPLGVPDILLMVSDAVVVF
DNLAGKMHAIVLVDPAEEQAFEQGQARLQGLLETLRQPITPRRGLDLSGPQAAEPEFR
SSYTREDYENAVGRIKEYILAGDCMQVVPSQRMSIDFKAAPIDLYRALRCFNPTPYMY
FFNFGDFHVVGSSPEVLVRVEDNLVTVRPIAGTRPRGATEEADRALEDDLLSDDKEIA
EHLMLIDLGRNDVGRVSSTGSVRLTEKMVIERYSNVMHIVSNVAGQLREGLTAMDALR
AILPAGTLSGAPKIRAMEIIDELEPVKRGVYGGAVGYFAWNGNMDTAIAIRTAVINDG
ELHVQAGGGIVADSVPALEWEETINKRRAMFRAVALAEQTSAK"

CDS 4366. .4452
/note="orf"
/codon_start=1
/transl_table=11
/protein_id="AAA80552.1"
/db_xref="GI:151630"
/translation="MSETYPVLIVSKLLMRVIKAHARWRWRA"

gene 4589. .5185

```
                     /gene="trpG"
     CDS             4589. .5185
                     /gene="trpG"
                     /codon_start=1
                     /transl_table=11
                     /product="anthranilate synthetase beta-subunit"
                     /protein_id="AAA80553.1"
                     /db_xref="GI:151631"
                     /translation="MLLMMIDNYDSFTYNVVQYLGELGAEVKVIRNDEMTIAQIEALN
                     PERIVVSPGPCTPSEAGVSIEAILHFAGKLPILGVCLGHQSIGQAFGGDVVRARQVMH
                     GKTSPVHHRDLGVFTGLNNPLTVTRYHSLVVKRETLPDCLEVTAWTAHEDGSVDEIMG
                     LRHKTLNIEGVQFHPESILTEQGHELFANFLKQTGGRR"
     gene            5195. .6244
                     /gene="trpD"
     CDS             5195. .6244
                     /gene="trpD"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoribosyl transferase"
                     /protein_id="AAA80554.1"
                     /db_xref="GI:1052829"
                     /translation="MDIKSALSRIVGQLDLTTEEMRDVMRQIMTGQCTEAQIGAFLMG
                     MRMKSESIDEIVGAVSVMRELADKVELKSLDGVVDIVGTGGDGANIFNVSTASSFVLA
                     AAGCTVAKHGNRAVSGKSGSADLLEAAGIYLNLTPTQVARCIDSLGIGFMFAQSHHSA
                     MKHAAGPRRDLGLRTLFNMLGPLTNPAGVKHQVVGVFAQTLCRPLAEVLQRLGSKHVL
                     VVHSKDGLDEFSLAAPTFVAELKNGEITEYWVEPEDLGMKSQSLHGLAVENPQASLEL
                     IRDALGRRKTENGQKAAEMIVLNAGAALYAADHAMTLAQGVELAHDVLHTGLAWEKLQ
                     ELGAFTAVFKVENEA"
     gene            6241. .7074
                     /gene="trpC"
     CDS             6241. .7074
                     /gene="trpC"
                     /codon_start=1
                     /transl_table=11
                     /product="indoleglycerol phosphate synthetase"
                     /protein_id="AAA80555.1"
                     /db_xref="GI:1052830"
                     /translation="MSVPTVLERIIARKFQEVAERSARVSLAELERLAKTADAPRGFA
                     NALIEQAKRKQPAVIAEIKKASPSKGVIREHFVPAEIAVSYEKGGATCLSVLTDVDYF
                     QGADEYLQQARAAVSLPVIRKDFMVDPYQIVEARALGADCVLLIVSALDDVKMAELAA
                     TAKDVGLDVLVEVHDGDELERALKTLDTPLVGVNNRNLHTFEVSLETTLDLLPRIPRD
                     RLAITESGILNRADVELMAINEVYSFLVGEAFMRAEQPGLELQRLFFPEQVKKTVQPL
                     D"
ORIGIN

  Query Match           30.5%;  Score 18;  DB 1;  Length 7419;
  Best Local Similarity  100.0%;  Pred. No. 23;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 GACACCCTGCCGCAGAAA 36
              ||||||||||||||||||
Db        388 GACACCCTGCCGCAGAAA 405


RESULT 15
AC004625/c
LOCUS       AC004625               42499 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone T26J13 map CIC11C08,
            complete sequence.
ACCESSION   AC004625
VERSION     AC004625.3  GI:20197200
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 42499)
  AUTHORS   Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
            Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42499)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 42499)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598427.
FEATURES             Location/Qualifiers
     source          1. .42499
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="CIC11C08"
                     /clone="T26J13"
     misc_feature    complement(1. .5112)
                     /note="overlap with BAC clone F13H10 (AC005662:1. .5112)."
     gene            2683. .3649
                     /gene="At2g41430"
                     /note="synonym: T26J13.2; identical to GB D30719;
                     supported by full length cDNA: Ceres: 31388"
     mRNA            join(2683. .2727,2957. .3312,3400. .3649)
                     /gene="At2g41430"
     CDS             join(2973. .3312,3400. .3551)
                     /gene="At2g41430"
                     /codon_start=1
                     /product="ERD15 protein"
                     /protein_id="AAC23728.1"
                     /db_xref="GI:3241941"
                     /translation="MAMVSGRRSTLNPDAPLFIPAAVRQVEDFSPEWWQLVTTSTWYP
                     DYWISQQQQGADGFYDNGENENGGGHIDVADLLPESFDFDDMEDFFDTDAAEFDQGFD
                     GRMYYQAPSEFGFGKNGEMVKKSSGNRSPRSIVEPAKYAEKPAKWGNQRVAAAPRNIH
                     QPR"
     gene            4440. .10934
                     /gene="At2g41440"
                     /note="synonym: T26J13.3; predicted by genscan"
     mRNA            join(<4440. .4488,4522. .4676,4989. .5123,5491. .5528,
                     5605. .5701,5791. .5832,6102. .6175,6573. .6615,6694. .6773,
                     6869. .7010,7136. .7268,8763. .8916,9258. .9300,9378. .9418,
                     9516. .9615,9739. .9789,10210. .10289,10327. .10421,
                     10852. .>10934)
                     /gene="At2g41440"
     CDS             join(4440. .4488,4522. .4676,4989. .5123,5491. .5528,
                     5605. .5701,5791. .5832,6102. .6175,6573. .6615,6694. .6773,
                     6869. .7010,7136. .7268,8763. .8916,9258. .9300,9378. .9418,
                     9516. .9615,9739. .9789,10210. .10289,10327. .10421,
                     10852. .10934)
                     /gene="At2g41440"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAC23729.1"
                     /db_xref="GI:3241942"
                     /translation="MAPKRASPMQSSLYSIENSDQSNESSCVLIRFLQVITRLQWMVR
                     ISPVNPDPKAFKSNFPETTQSKSQWKKRKIGFRSIRFDGERGRSDEEEAKRSNEQDER
                     ALQFLRILQQQLRKEVERLRLLIKRMTGKEDIVSFTELLALQSHLMDVRSIVLEQKKE
                     VELEEAERPHKQKKEAEHDADGMRRGVLVPCKFSASSLGERQEDPNGEAPLLQMQREF
                     EQRSSESLQSEFDKLWLFNERMNGRELEGMTSFDLSILHIQILRALRALIDQKLGPRQ
                     EHRAWQRGELSPGDVRRIMTRMLTLPCNVDFFSFQESSEDSIILISGRACRRPDINFG
                     FPSIRLEISRMGRGRIKMKLKGRRRDLTKKKVRELCDYCGFSSGIIRYDEQRPHERDH
                     EAVSMENELKRLRLLLTRRMTCKDLDGLTFPELLSLKSHLQTALLIVKDQTKKIEQLLK
                     EDDGWMLIRSQLQLQFERRNAEPMMIEYGRLWLLKEVDVSFTLFLMVCFFFLLFVKTD
                     RRMNGRQLDGVGVRDRFFFFGYRVYSDRRLRTIRVF"
     repeat_region   8133. .8176
                     /rpt_family="AT_rich"
     repeat_region   8261. .8300
                     /rpt_family="AT_rich"
     repeat_region   complement(8423. .8563)
                     /rpt_family="AT_rich"
     repeat_region   complement(8590. .8611)
```

```
                     /rpt_family="(TA)n"
repeat_region        9707. .9729
                     /rpt_family="AT_rich"
gene                 11427. .18324
                     /gene="At2g41450"
                     /note="synonym: T26J13.4; predicted by genscan"
mRNA                 join(<11427. .11475,11562. .11585,11869. .11963,12598. .12690,
                     12937. .13047,13182. .13455,13575. .13993,14073. .14182,
                     14281. .14339,14419. .14606,14686. .14706,15252. .15372,
                     15585. .15755,16207. .16481,16572. .16721,16826. .16878,
                     17126. .17234,17326. .17460,17549. .17593,17684. .17871,
                     17960. .18101,18181. .>18324)
                     /gene="At2g41450"
CDS                  join(11427. .11475,11562. .11585,11869. .11963,12598. .12690,
                     12937. .13047,13182. .13455,13575. .13993,14073. .14182,
                     14281. .14339,14419. .14606,14686. .14706,15252. .15372,
                     15585. .15755,16207. .16481,16572. .16721,16826. .16878,
                     17126. .17234,17326. .17460,17549. .17593,17684. .17871,
                     17960. .18101,18181. .18324)
                     /gene="At2g41450"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC23730.1"
                     /db_xref="GI:3241943"
                     /translation="MAPKRASPMQSSLYSIGDYEVAVDEKDKIVSDYPKEDMFLLLNP
                     RDEDDFTKSHLQILAAITYQIVPADTQYAEIPLAAVTYTHQKKGFGKLVYEELMKRLH
                     SVGIRTIYCWADKESEGFWLKQGFIKLAEVDHKGKAKGLHIKSNIRKALCFPGGSTLM
                     LSHLNKEFISNLEISSSWKYQCEGSPLSARNNSTGPVTGDSVKLRESFGESVYLDCIS
                     GIRSPMDSITGKENNNVISDQATTADSETKCSTPGLKRSWEASLSSLQSKRIRANNNN
                     NSEIAKTDLAQSSAKQSKDGNSSQVDITKDSLPTICKRNDVEQCRMATGIDMEARFNG
                     QHYRILLMDICDENKRAWLTEVIRKLGGTVTLDGTTSTHIVTGKVRKTLNLCTALCSG
                     AWIVSPSWLKESVREGRFANEASHILHDEDYQLKYDTDLKSTVLRAKARPNSLLKGYD
                     ICVGPNIELPIKTSSAIIKSAGGNVISGVNKGKAKTFRIALSLAHGSVRVSCHFHHYR
                     SLKTFFQLILFVNQFVDFKKPISGVDAEMKIVRYLLQNSAILKKLTIRLVKRSWKAEC
                     FLEDLCCIQRGSNSSYSRKFLLITFLPLSLACLAFLLQWRSGVNDSVTQWFDDNYPFP
                     GMATVSEKRSLRSDSSCVSLLGQSRTQAFPYLRDLKLDHKPDLKPRICITTSTSAGLE
                     QTLPWIFYHKVIGVETFYLFVEGTAASPNVSRVLETIPGVNVIYRTRELEEEQAKSRI
                     WNETWLEKFFYKPCNYELFVKQNLNMEMAITMARDAGMDWILHLDTDELVHPSGTREY
                     SLRNLLRDVPADVDEVIFTNYESSVERDDIKEPFTEVSMFKKNFKHLPREVYYGNYKE
                     ATRGNPNYFLTYANGKSAARIQDHLRPNGAHRWHNYKTYPNIKELDEAAILHYTYSKF
                     SDLTSRRDRCGCKPTKKDVKRCFMLDFDRAAFIIASTSTSEEMLQWYRERVVWTDDNL
                     ILKLLRKGILTRIYAPMVS"
gene                 18968. .22165
                     /gene="At2g41460"
                     /note="synonym: T26J13.5; identical to GB:X76912"
mRNA                 join(<18968. .19020,19356. .19536,19634. .19728,19855. .19934,
                     20201. .20419,20513. .20835,20937. .21047,21129. .21266,
                     21363. .21410,21494. .21574,21652. .21811,21885. .22165)
                     /gene="At2g41460"
CDS                  join(18968. .19020,19356. .19536,19634. .19728,19855. .19934,
                     20201. .20419,20513. .20835,20937. .21047,21129. .21266,
                     21363. .21410,21494. .21574,21652. .21811,21885. .22006)
                     /gene="At2g41460"
                     /codon_start=1
                     /product="DNA-(apurinic or apyrimidinic site) lyase (ARP)"
                     /protein_id="AAC23731.1"
                     /db_xref="GI:3241944"
                     /translation="MNNVLQFGLQSSAIYVAKFLVVPLRSLRVGSSFVGVGVGTRSFN
                     KRLMSNATAFSINNSKRKELKIPGAAIDQNCHQMGSDTDRDEMGTLQDDRKEIEAMTV
                     QELRSTLRKLGVPVKGRKQELISTLRLHMDSNLPDQKETSSSTRSDSVTIKRKISNRE
                     EPTEDECTNSEAYDIEHGEKRVKQSTEKNLKAKVSAKAIAKEQKSLMRTGKQQIQSKE
                     ETSSTISSELLKTEEIISSPSQSEPWTVLAHKKPQKDWKAYNPKTMRPPPLPEGTKCV
                     KVMTWNVNGLRGLLKFESFSALQLAQRENFDILCLQETKLQVKDVEEIKKTLIDGYDH
                     SFWSCSVSKLGYSGTAIISRIKPLSVRYGTGLSGHDTEGRIVTAEFDSFYLINTYVPN
                     SGDGLKRLSYRIEEWDRTLSNHIKELEKSKPVVLTGDLNCAHEEIDIFNPAGNKRSAG
                     FTIEERQSFGANLLDKGFVDTFRKQHPGVVGYTYWGYRHGGRKTNKGWRLDYFLVSQS
                     IAANVHDSYILPDINGSDHCPIGLILKL"
repeat_region        complement(24160. .24183)
                     /rpt_family="(A)n"
gene                 complement(25382. .29598)
                     /gene="At2g41470"
                     /note="synonym: T26J13.6"
mRNA                 complement(join(<25382. .25532,26138. .26199,26283. .26315,
                     26420. .26516,26603. .26640,26955. .27081,27462. .27515,
                     28500. .28798,29256. .29363,29438. .>29598))
                     /gene="At2g41470"
CDS                  complement(join(25382. .25532,26138. .26199,26283. .26315,
                     26420. .26516,26603. .26640,26955. .27081,27462. .27515,
                     28500. .28798,29256. .29363,29438. .29566))
                     /gene="At2g41470"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAC23732.1"
                     /db_xref="GI:3241945"
                     /translation="MLRLAIPLFLFALCSFTLFSSARSFITTKPLPIDSFIPKPKLEN
                     AAACSYTVIIKTSCSSVSYTRDKISISFGDVYGNEVYVKRLDDPSSRTFEKCSSDTYK
                     ISGPCMRDVCYLYLLRQGSDGWKPENVKIYGSSIRSVTFYYNLFLPNSVWYGFNVCNG
                     IGNTKSSQPISTTSSVAAINLRRFHNSYQRKSMPETSSFNIHKRWDKKKHIYSHIGFI
                     EELANRFEEEDVQLLRVLQQQLREKVERLRFLIKRMTGKDDDAHFTELQALESSRLKDV
                     SRTVLDQKEKMIKLEEDERKKEAQNESDGTRRGVLVPCNRRMNGRELEGMSSRDLLSL
                     DNQILNALLGLLVQMNRPREEQVAARQREKV"
repeat_region        complement(26656. .26681)
                     /rpt_family="(TA)n"
repeat_region        29864. .29901
                     /rpt_family="AT_rich"
gene                 complement(30223. .32014)
                     /gene="At2g41480"
                     /note="synonym: T26J13.7"
mRNA                 complement(join(<30223. .30644,31172. .31334,31436. .31615,
                     31706. .>32014))
                     /gene="At2g41480"
CDS                  complement(join(30223. .30644,31172. .31334,31436. .31615,
                     31706. .32014))
                     /gene="At2g41480"
                     /codon_start=1
                     /product="putative peroxidase"
                     /protein_id="AAC23733.1"
                     /db_xref="GI:3241946"
                     /translation="MYVCEYKYKLVLEHHCMLINHTAQIRVRKMGVYLGKYCYIMIIM
                     LVLVLGKEVRSQLLKNGYYSTSCPKAESIVRSTVESHFDSDPTISPGLLRLHFHDCFV
```

```
 Query Match             30.5%;  Score 18;  DB 8;  Length 42499;
 Best Local Similarity   100.0%;  Pred. No. 23;
 Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 GCAGAAACCGCGTGAATC 47
              ||||||||||||||||||
Db       1361 GCAGAAACCGCGTGAATC 1344


Search completed: December 22, 2004, 23:36:38
Job time : 841.543 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:14:28 ; Search time 208.235 Seconds
(without alignments)
1487.336 Million cell updates/sec

Title: US-09-898-616A-3
Perfect score: 59
Sequence: 1 cagctgaagaaactggttga.....cgtgaatccatcataaactg 59

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	59	100.0	59	9	ABZ58372	Abz58372 Human ute
	2	59	100.0	59	12	ADL27628	Adl27628 Recombina
c	3	37	62.7	59	9	ABZ58375	Abz58375 Human ute
c	4	37	62.7	59	12	ADL27631	Adl27631 Recombina
	5	18	30.5	215	1	AAN90906	Aan90906 Synthetic
c	6	17	28.8	195	6	ABK76145	Abk76145 Bacillus
c	7	17	28.8	993	10	ADC24050	Adc24050 DNA seque
c	8	17	28.8	993	12	ADH36151	Adh36151 Chemical
c	9	17	28.8	993	12	ADG93852	Adg93852 Nitrilase
c	10	17	28.8	993	12	ADI62449	Adi62449 DNA encod
c	11	17	28.8	993	12	ADI64570	Adi64570 DNA encod
	12	17	28.8	3530	4	ABL08480	Abl08480 Drosophil
c	13	17	28.8	5568	10	ABX08140	Abx08140 S. pneumo
c	14	17	28.8	5568	12	ADM91810	Adm91810 S pneumon
c	15	17	28.8	14736	2	AAV52304	Aav52304 Streptoco
c	16	17	28.8	62598	10	ABS56454_21	Continuation (22 o
c	17	16	27.1	60	9	ABZ58374	Abz58374 Human ute
c	18	16	27.1	60	12	ADL27630	Adl27630 Recombina
	19	16	27.1	311	4	AAL01419	Aal01419 Human rep
	20	16	27.1	377	8	ABX55377	Abx55377 Bovine ES
c	21	16	27.1	603	11	ABD04549	Abd04549 Pseudomon
c	22	16	27.1	639	10	ADB69621	Adb69621 C. neofor
c	23	16	27.1	659	3	AAF08058	Aaf08058 Fusarium
	24	16	27.1	667	6	ABQ75940	Abq75940 CHD activ
c	25	16	27.1	743	10	ADB69260	Adb69260 C. neofor
	26	16	27.1	933	11	ABD02206	Abd02206 Pseudomon
c	27	16	27.1	1346	5	AAS73783	Aas73783 DNA encod
	28	16	27.1	1769	6	ABN98233	Abn98233 Arabidops
c	29	16	27.1	2142	11	ABD02265	Abd02265 Pseudomon
	30	16	27.1	2658	5	AAS90260	Aas90260 DNA encod
c	31	16	27.1	2658	5	AAS94408	Aas94408 DNA encod
c	32	16	27.1	2658	5	AAS73288	Aas73288 DNA encod
c	33	16	27.1	2659	5	AAS88462	Aas88462 DNA encod
c	34	16	27.1	2743	10	ADB68899	Adb68899 C. neofor
c	35	16	27.1	2874	8	ACA18781	Aca18781 Prokaryot
	36	16	27.1	3056	3	AAC42758	Aac42758 Arabidops
	37	16	27.1	3108	4	ABL24046	Abl24046 Drosophil
c	38	16	27.1	3163	10	ADE25034	Ade25034 Plant gro
	39	16	27.1	4221	12	ADO57346	Ado57346 DNA encod
c	40	16	27.1	5603	4	ABL12539	Abl12539 Drosophil
c	41	16	27.1	6564	3	AAC55314	Aac55314 Human act
	42	16	27.1	10327	4	ABL12538	Abl12538 Drosophil
c	43	16	27.1	11204	3	AAC55339	Aac55339 Human act
c	44	16	27.1	11204	6	ABS73286	Abs73286 DNA encod
	45	16	27.1	21469	4	AAK89568	Aak89568 Human dig

ALIGNMENTS

RESULT 1
ABZ58372
ID ABZ58372 standard; DNA; 59 BP.
XX
AC ABZ58372;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 3.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Claim 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 3, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and

CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 59 BP; 19 A; 17 C; 13 G; 10 T; 0 U; 0 Other;

Query Match 100.0%; Score 59; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59

RESULT 2
ADL27628
ID ADL27628 standard; DNA; 59 BP.
XX
AC ADL27628;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, coding oligonucleotide #3.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Claim 1; SEQ ID NO 3; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a coding strand oligonucleotide used to
CC construct the synthetic rhUG gene.
XX
SQ Sequence 59 BP; 19 A; 17 C; 13 G; 10 T; 0 U; 0 Other;

Query Match 100.0%; Score 59; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59

RESULT 3
ABZ58375/c
ID ABZ58375 standard; DNA; 59 BP.
XX
AC ABZ58375;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 6.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 6, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the

```
CC   complementary strand. The gene was assembled by annealing and ligation of
CC   the oligonucleotides. Because mature native hUG has glutamic acid at its
CC   N-terminus, an initiator methionine was added to the N-terminus, and
CC   codon usage was optimised for expression in bacteria. In an example from
CC   the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC   ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC   coli strain CG12. The invention relates generally to the production of
CC   recombinant hUG by bacterial expression, protein purification and scaled-
CC   up production according to current good manufacturing practices. The
CC   recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC   conditions, such as neonatal respiratory distress syndrome and
CC   bronchopulmonary dysplasia. It may also be used to treat conditions
CC   associated with elevated phospholipase A2 levels such as pancreatitis,
CC   acute renal failure, rheumatoid arthritis and asthma
XX
SQ   Sequence 59 BP; 9 A; 17 C; 15 G; 18 T; 0 U; 0 Other;

  Query Match             62.7%;  Score 37;  DB 9;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-10;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 4
ADL27631/c
ID   ADL27631 standard; DNA; 59 BP.
XX
AC   ADL27631;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #2.
XX
KW   Human; ss; recombinant human uteroglobin; rhUG;
KW   bacterial expression system; rhUG master cell bank;
KW   rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW   fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2003207795-A1.
XX
PD   06-NOV-2003.
XX
PF   02-JUL-2002; 2002US-00187498.
XX
PR   28-MAY-1997;   97US-00864357.
PR   02-JUL-2001; 2001US-00898616.
XX
PA   (PILO/) PILON A L.
PA   (WELC/) WELCH R W.
XX
PI   Pilon AL,  Welch RW;
XX
DR   WPI; 2004-051527/05.
XX
PT   Bacterial expression system for production of recombinant human
PT   uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT   for human uteroglobin.
XX
PS   Example 1; SEQ ID NO 6; 64pp; English.
XX
CC   The invention relates to a bacterial expression system for the production
CC   of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC   human cDNA sequence which codes for human UG, constructed from the
CC   oligonucleotides appearing as ADL27626-ADL27629, and which further
CC   comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC   are producing an rhUG master cell bank (comprising inoculating a suitable
CC   incubating broth with an aliquot portion of a rhUG research seed bank to
CC   form a bacterial culture, incubating the bacterial culture, adding a
CC   cryopreservative to the bacterial culture to form a cryopreserved
CC   solution, transferring a portion of the cryopreserved solution to a
CC   cryovial and storing the cryovial at a temperature below -60 degrees C),
CC   expressing rhUG (comprising providing a production seed cell bank culture
CC   comprising an expression vector capable of expressing rhUG, inoculating a
CC   broth medium with the production seed cell bank culture to form an
CC   inoculum, incubating the bacterial culture formed in step (b),
CC   inoculating a large scale fermenter with the inoculum formed from the
CC   step (c) to form a fermentation culture, incubating the fermentation
CC   culture within the large scale fermenter, adding an induction agent to
CC   the fermentation culture to induce the expression of rhUG and harvesting
CC   the above fermentation culture), purifying rhUG, determining the potency
CC   of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC   from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC   rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC   purity of rhUG, and a pharmaceutical composition comprising a purified
CC   rhUG and a carrier or diluent. The bacterial expression system is useful
CC   for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC   drug substance. rhUG is safe to administer to a patient in respiratory
CC   distress. The rhUG is useful for treating inflammation and fibrotic
CC   diseases. The present sequence is a non-coding strand oligonucleotide
CC   used to construct the synthetic rhUG gene.
XX
SQ   Sequence 59 BP; 9 A; 17 C; 15 G; 18 T; 0 U; 0 Other;

  Query Match             62.7%;  Score 37;  DB 12;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-10;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 5
AAN90906
ID   AAN90906 standard; DNA; 215 BP.
XX
AC   AAN90906;
XX
DT   25-MAR-2003  (revised)
DT   26-JAN-1990  (first entry)
XX
DE   Synthetic human IGF-II.
XX
KW   Synthetic IGF-II; fusion protein.
XX
OS   Homo sapiens.
XX
PN   EP333691-A.
XX
PD   20-SEP-1989.
XX
PF   16-MAR-1989;   89EP-00850091.
XX
PR   17-MAR-1988;   88SE-00000981.
XX
PA   (CEMU-) CEMU BIOTEKNIK AB.
PA   (SEMB-) SEM BIOTECHNIK AB.
XX
PI   Nygren PA,  Abrahmsen L,  Uhlen M;
XX
DR   WPI; 1989-272436/38.
DR   P-PSDB; AAP91389.
XX
PT   New recombinant fusion protein comprising desired protein - flanked by
PT   IgG-binding domain of staphylococcal protein A and albumin-binding domain
PT   of streptococcal protein G.
XX
PS   Disclosure; Fig 1; 10pp; English.
XX
CC   Synthetic IGF-II was constructed from 22 oligonucleotides. The gene was
```

CC designed with EcoRI and HindIII 5' and 3' cohesive ends resp. to
CC facilitate cloning. An N-terminal methionine codon was included to allow
CC cleavage of fusion proteins with CNBr to generate native IGF-II. A double
CC TAA stop codon was incorporated at the 3' end. the gene was used to
CC construct a fusion vector with DNA coding for an IgG-binding domain from
CC staphylococcal protein A and an albumin-binding part of streptococcal
CC protein G. (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 215 BP; 35 A; 55 C; 53 G; 72 T; 0 U; 0 Other;

Query Match 30.5%; Score 18; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AACTGGTTGACACCCTGC 28
||||||||||||||||||
Db 44 AACTGGTTGACACCCTGC 61

RESULT 6
ABK76145/c
ID ABK76145 standard; DNA; 195 BP.
XX
AC ABK76145;
XX
DT 13-AUG-2002 (first entry)
XX
DE Bacillus licheniformis genomic sequence tag (GST) #3436.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX
DR WPI; 2002-416684/44.
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX
PS Claim 4; SEQ ID NO 3436; 200pp; English.
XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC -up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 195 BP; 34 A; 48 C; 56 G; 57 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 6; Length 195;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 CACCCTGCCGCAGAAAC 37
|||||||||||||||||
Db 102 CACCCTGCCGCAGAAAC 86

RESULT 7
ADC24050/c
ID ADC24050 standard; DNA; 993 BP.
XX
AC ADC24050;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA sequence (SeqID 317) encoding a nitrilase enzyme.
XX
KW gene; ds; nitrilase; nitrile; cyanohyrdin; ammonia; biocatalyst;
KW enantiomer; chiral medicine.
XX
OS Unidentified.
XX
PN WO2003000840-A2.
XX
PD 03-JAN-2003.
XX
PF 15-MAY-2002; 2002WO-US015983.
XX
PR 21-JUN-2001; 2001US-0300189P.
PR 30-JUL-2001; 2001US-0309006P.
PR 22-JAN-2002; 2002US-0351336P.
XX
PA (DIVE-) DIVERSA CORP.
PA (MADD/) MADDEN D.
XX
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;
PI Short JM, Burk M;
XX
DR WPI; 2003-201417/19.
DR P-PSDB; ADC24051.
XX
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl
PT lactic acid derivative and for producing pharmaceutical composition, and
PT food additive.
XX
PS Claim 1; SEQ ID NO 317; 560pp; English.
XX
CC This invention relates to nitrilases and the nucleic acids that encode
CC these enzymes thereof. Specifically, it refers to polypeptides that
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse
CC nitriles or cyanohyrdins into their corresponding carboxylic acids and
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in
CC the synthesis of enantiomerically pure aromatic and aliphatic amino
CC acids, as well as hydroxy acids, which are important for the development
CC of chiral medicines. Furthermore, the present invention describes
CC nitrilases, isolated from mesophilic microorganisms, that have improved
CC activity and stability at increased pH and temperature. They are also
CC inexpensive, efficient catalysts, have broad substrate specificity and
CC are capable of chiral differentiation. This polynucleotide is a DNA
CC sequence that encodes a nitrilase enzyme of the invention.
XX

SQ Sequence 993 BP; 188 A; 330 C; 305 G; 170 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 10; Length 993;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 CCGCGTGAATCCATCAT 53
||||||||||||||||
Db 860 CCGCGTGAATCCATCAT 844

RESULT 8
ADH36151/c
ID ADH36151 standard; DNA; 993 BP.
XX
AC ADH36151;
XX
DT 11-MAR-2004 (first entry)
XX
DE Chemical process monitoring-related nitrilase gene sequence SeqID317.
XX
KW chemical process monitoring; biochemical process monitoring; cyanide;
KW high throughput system; gene; ds.
XX
OS Unidentified.
XX
PN WO2003098187-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003WO-US015639.
XX
PR 15-MAY-2002; 2002US-0380737P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI Mcquaid J, Stege J;
XX
DR WPI; 2004-142708/14.
DR P-PSDB; ADH36152.
XX
PT Monitoring a chemical or biochemical process comprises providing a
PT reactant comprising a cyanide or a material that can be converted to
PT cyanide or a reactant that generates a cyanide or a material that can be
PT converted to cyanide.
XX
PS Claim 74; SEQ ID NO 317; 277pp; English.
XX
CC This invention relates to a novel method of monitoring chemical or
CC biochemical processes. The method involves providing a reactant
CC comprising cyanide (or a material that can be converted to a cyanide)
CC that generates as a reaction product cyanide or a material that can be
CC converted to cyanide and measuring the concentration of produced cyanide.
CC The method is useful for monitoring a chemical or biochemical process.
CC The method is effective for high throughput systems and is sufficiently
CC sensitive to detect a small amount of product. The present sequence is
CC that of a gene which encodes a nitrilase enzyme which can be used in the
CC method of the invention.
XX
SQ Sequence 993 BP; 188 A; 330 C; 305 G; 170 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 12; Length 993;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 CCGCGTGAATCCATCAT 53
||||||||||||||||
Db 860 CCGCGTGAATCCATCAT 844

RESULT 9
ADG93852/c
ID ADG93852 standard; DNA; 993 BP.
XX
AC ADG93852;
XX
DT 11-MAR-2004 (first entry)
XX
DE Nitrilase enzyme gene sequence SeqID317.
XX
KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;
KW enantioselective transformation; gene; ds.
XX
OS Unidentified.
XX
PN WO2003097810-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003WO-US015712.
XX
PR 15-MAY-2002; 2002US-00146772.
PR 09-SEP-2002; 2002US-00241742.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;
XX
DR WPI; 2004-090637/09.
DR P-PSDB; ADG93853.
XX
PT New isolated or recombinant nucleic acid encoding a polypeptide having
PT nitrilase activity, useful for screening enantioselective transformation.
XX
PS Claim 44; SEQ ID NO 317; 295pp; English.
XX
CC This invention is related to a novel isolated or recombinant nucleic acid
CC encoding a protein having nitrilase activity. Nitrilase's are capable of
CC converting nitrile's directly to carboxylic acids and have great
CC potential for use in industrial chemical processes. The isolated
CC nitrilase proteins of the invention have increased activity and stability
CC at increased pH and temperature when compared to those conventionally
CC used. In addition, the nucleic acid of the invention is useful for
CC screening enantioselective transformation. The present sequence is that
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.
XX
SQ Sequence 993 BP; 188 A; 330 C; 305 G; 170 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 12; Length 993;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 CCGCGTGAATCCATCAT 53
||||||||||||||||
Db 860 CCGCGTGAATCCATCAT 844

RESULT 10
ADI62449/c
ID ADI62449 standard; DNA; 993 BP.
XX
AC ADI62449;
XX
DT 22-APR-2004 (first entry)
XX
DE DNA encoding nitrilase polypeptide #159.
XX
KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;
KW antilipaemic; gene; ds.
XX
OS Unidentified.

XX
PN WO2003106415-A2.
XX
PD 24-DEC-2003.
XX
PF 13-JUN-2003; 2003WO-US018840.
XX
PR 13-JUN-2002; 2002US-0389317P.
PR 28-JUN-2002; 2002US-0392944P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Burk M, Desantis G, Morgan B, Zhu Z;
XX
DR WPI; 2004-090821/09.
DR P-PSDB; ADI62450.
XX
PT Preparation of atorvastatin comprises catalytic conversion of 3-
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
PT acid and forming atorvastatin.
XX
PS Claim 46; SEQ ID NO 317; 253pp; English.
XX
CC The present invention relates to a method for preparating an atorvastatin
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
CC hydroxybutyric acid with a polypeptide having nitrilase activity,
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
CC method involves whole cell processes, cell lysate process, "one pot"
CC processes, and "multi-pot" processes using a variety of parameters.
CC Atorvastatin is used, in conjunction with dietary restriction, in the
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present
CC sequence encodes a nitrilase polypeptide obtained from an environmental
CC sample.
XX
SQ Sequence 993 BP; 188 A; 330 C; 305 G; 170 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 12; Length 993;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 CCGCGTGAATCCATCAT 53
||||||||||||||||
Db 860 CCGCGTGAATCCATCAT 844

RESULT 11
ADI64570/c
ID ADI64570 standard; DNA; 993 BP.
XX
AC ADI64570;
XX
DT 22-APR-2004 (first entry)
XX
DE DNA encoding nitrilase seq id 159.
XX
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
KW carboxylic acid; cyanohydrin moiety hydrolysis;
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
KW (R)-phenyl lactic acid derivative; % enantiomeric excess;
KW % diasteromeric excess; food additive; drug intermediate; ds; nitrilase;
KW gene.
XX
OS Unidentified.
XX
PN US2004014195-A1.
XX
PD 22-JAN-2004.
XX
PF 15-MAY-2003; 2003US-00440523.
XX
PR 29-DEC-1999; 99US-0173609P.
PR 07-DEC-2000; 2000US-0254414P.
PR 28-DEC-2000; 2000US-00751299.
PR 21-JUN-2001; 2001US-0300189P.
PR 30-JUL-2001; 2001US-0309006P.
PR 22-JAN-2002; 2002US-0351336P.
PR 15-MAY-2002; 2002US-00146772.
PR 09-SEP-2002; 2002US-00241742.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX
DR WPI; 2004-121569/12.
DR P-PSDB; ADI64571.
XX
PT Novel isolated or recombinant polypeptide having nitrilase activity,
PT useful in production of food additives.
XX
PS Claim 1; SEQ ID NO 317; 105pp; English.
XX
CC The invention describes an isolated or recombinant polypeptide (I)
CC comprising amino acids having a sequence at least 50 % identical to a
CC sequence (S1) available in electronic form (EC) from the following web
CC site ftp.seqdata.uspto.gov/sequence.html?DocID=2004014195, or its
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
CC and for identifying a modified compound. The inventive method is useful
CC for monitoring or determining % enantiomeric excess or % diasteromeric
CC excess. (I) is useful in the production of food additives and drug
CC intermediates. This sequence encodes a nitrilase of the invention.
XX
SQ Sequence 993 BP; 188 A; 330 C; 305 G; 170 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 12; Length 993;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 CCGCGTGAATCCATCAT 53
||||||||||||||||
Db 860 CCGCGTGAATCCATCAT 844

RESULT 12
ABL08480
ID ABL08480 standard; cDNA; 3530 BP.
XX
AC ABL08480;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19922.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.

```
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB64377.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 19922; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3530 BP; 1121 A; 634 C; 690 G; 1085 T; 0 U; 0 Other;

  Query Match             28.8%;  Score 17;  DB 4;  Length 3530;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CTGAAGAAACTGGTTGA 20
              |||||||||||||||||
Db       1786 CTGAAGAAACTGGTTGA 1802


RESULT 13
ABX08140/c
ID   ABX08140 standard; DNA; 5568 BP.
XX
AC   ABX08140;
XX
DT   27-OCT-2003  (revised)
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain coding region #2428.
XX
KW   Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW   ear infection; antiinflammatory; antibacterial; immunostimulant;
KW   auditory; respiratory; gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae; type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB002163.
XX
PR   27-MAR-2001; 2001GB-00007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   P-PSDB; ABU02847.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT   ear infection.
XX
PS   Claim 6; SEQ ID NO 4855; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the proteins,
CC   treating a patient by administering the protein, DNA or antibody (in a
CC   composition), a kit comprising first and second primers, which are the
CC   nucleic acid cited above or fragments between nucleotides 8-100 of a
CC   sequence not defined in the specification, for amplifying a target
CC   sequence contained within a Streptococcus nucleic acid sequence, where
CC   the first primer is substantially complementary to the target sequence
CC   and the second primer is substantially complementary to the complement of
CC   the target sequence, and where the parts of the primers having
CC   substantial complementarity define the termini of the target sequence to
CC   be amplified, assay comprising contacting a test compound with the
CC   protein, and determining whether the test compound binds to the protein
CC   and a Streptococcus pneumoniae bacterium, where one or more genes
CC   encoding the proteins has been rendered inactive. The proteins, nucleic
CC   acid molecules, antibody and compositions are useful as medicaments for
CC   treating or preventing a disease or infection due to streptococcus
CC   bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC   media or ear infection. They are also useful in developing vaccines,
CC   diagnostics and antibiotics. The methods are useful for identifying
CC   immunodominant proteins. The present sequence is one of the 2489
CC   identified coding region from the genomic sequence. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 5568 BP; 2066 A; 873 C; 1182 G; 1447 T; 0 U; 0 Other;

  Query Match             28.8%;  Score 17;  DB 10;  Length 5568;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         42 TGAATCCATCATAAACT 58
              |||||||||||||||||
Db       4309 TGAATCCATCATAAACT 4293


RESULT 14
ADM91810/c
ID   ADM91810 standard; DNA; 5568 BP.
XX
AC   ADM91810;
XX
DT   03-JUN-2004  (first entry)
XX
DE   S pneumoniae antigenic protein-encoding gene sequence SeqID7.
XX
KW   antibacterial; gene therapy; Streptococcus pneumoniae infection;
KW   antigenic; gene; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO2004020609-A2.
XX
PD   11-MAR-2004.
XX
PF   02-SEP-2003; 2003WO-US027401.
XX
PR   30-AUG-2002; 2002US-0407082P.
XX
```

PA (TUFT) UNIV TUFTS.
XX
PI Camilli A, Hava DL;
XX
DR WPI; 2004-239189/22.
DR P-PSDB; ADM92047.
XX
PT New Streptococcus pneumoniae nucleic acid molecules, useful for
PT diagnosing, treating and preventing active infections of Streptococcus
PT pneumoniae.
XX
PS Claim 1; SEQ ID NO 7; 123pp; English.
XX
CC This invention relates to novel isolated Streptococcus pneumoniae nucleic
CC acid molecules and the antigenic polypeptides encoded by them. The
CC invention may be useful for the production of compounds with an
CC antibacterial activity or for gene therapy. The nucleic acid molecules,
CC compositions and methods disclosed are useful for treating Streptococcus
CC pneumoniae infection. The present sequence is that of an S pneumoniae
CC gene of the invention.
XX
SQ Sequence 5568 BP; 2066 A; 873 C; 1182 G; 1447 T; 0 U; 0 Other;

Query Match 28.8%; Score 17; DB 12; Length 5568;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 TGAATCCATCATAAACT 58
||||||||||||||||||
Db 4309 TGAATCCATCATAAACT 4293

RESULT 15
AAV52304/c
ID AAV52304 standard; DNA; 14736 BP.
XX
AC AAV52304;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:171.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019588.
XX
PR 31-OCT-1996; 96US-0029960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
PS Claim 1; Page 1085-1094; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridise to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae
XX
SQ Sequence 14736 BP; 4665 A; 2568 C; 3255 G; 4247 T; 0 U; 1 Other;

Query Match 28.8%; Score 17; DB 2; Length 14736;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 42 TGAATCCATCATAAACT 58
||||||||||||||||||
Db 13144 TGAATCCATCATAAACT 13128

Search completed: December 22, 2004, 22:44:17
Job time : 213.485 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 22:08:48 ; Search time 47.5204 Seconds
(without alignments)
882.496 Million cell updates/sec

Title: US-09-898-616A-3
Perfect score: 59
Sequence: 1 cagctgaagaaactggttga.....cgtgaatccatcataaactg 59

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		59	100.0	59	3	US-08-864-357F-8	Sequence 8, Appli
c 2		37	62.7	59	3	US-08-864-357F-11	Sequence 11, Appl
c 3		17	28.8	14736	4	US-08-961-527-171	Sequence 171, App
c 4		16	27.1	60	3	US-08-864-357F-10	Sequence 10, Appl
c 5		16	27.1	603	4	US-09-252-991A-3153	Sequence 3153, Ap
6		16	27.1	933	4	US-09-252-991A-810	Sequence 810, App
c 7		16	27.1	2142	4	US-09-252-991A-869	Sequence 869, App
8		15	25.4	309	4	US-09-489-039A-1471	Sequence 1471, Ap
9		15	25.4	342	4	US-09-252-991A-3202	Sequence 3202, Ap
c 10		15	25.4	369	4	US-09-489-039A-1498	Sequence 1498, Ap
c 11		15	25.4	484	4	US-09-270-767-7800	Sequence 7800, Ap
c 12		15	25.4	484	4	US-09-270-767-23082	Sequence 23082, A
c 13		15	25.4	516	4	US-09-270-767-1781	Sequence 1781, Ap
c 14		15	25.4	516	4	US-09-270-767-17063	Sequence 17063, A
c 15		15	25.4	525	4	US-09-489-039A-1476	Sequence 1476, Ap
c 16		15	25.4	768	4	US-09-270-767-29574	Sequence 29574, A
c 17		15	25.4	807	4	US-09-248-796A-4201	Sequence 4201, Ap
18		15	25.4	987	4	US-09-252-991A-3011	Sequence 3011, Ap
19		15	25.4	1092	4	US-09-543-681A-3173	Sequence 3173, Ap
c 20		15	25.4	1296	4	US-09-270-767-13572	Sequence 13572, A
c 21		15	25.4	1534	3	US-08-629-643A-4	Sequence 4, Appli
c 22		15	25.4	1534	3	US-09-155-884-4	Sequence 4, Appli
23		15	25.4	1605	4	US-09-807-802A-16	Sequence 16, Appl
24		15	25.4	1632	4	US-09-252-991A-3113	Sequence 3113, Ap
c 25		15	25.4	1743	4	US-09-252-991A-2863	Sequence 2863, Ap
26		15	25.4	1800	4	US-09-807-802A-14	Sequence 14, Appl
27		15	25.4	2211	4	US-09-807-802A-12	Sequence 12, Appl
28		15	25.4	3000	4	US-09-705-267A-18	Sequence 18, Appl
29		15	25.4	4718	4	US-09-807-802A-1	Sequence 1, Appli
c 30		15	25.4	6727	3	US-08-629-643A-5	Sequence 5, Appli
c 31		15	25.4	6727	3	US-09-280-799-1	Sequence 1, Appli
c 32		15	25.4	6727	3	US-09-155-884-5	Sequence 5, Appli
33		15	25.4	7447	4	US-10-216-870-11	Sequence 11, Appl
c 34		15	25.4	9347	4	US-10-204-708-35	Sequence 35, Appl
35		15	25.4	10480	4	US-09-732-615-13	Sequence 13, Appl
36		15	25.4	10480	4	US-10-273-051-13	Sequence 13, Appl
37		15	25.4	35100	2	US-08-770-379-18	Sequence 18, Appl
38		15	25.4	35100	3	US-08-757-669A-18	Sequence 18, Appl
39		15	25.4	35100	3	US-09-230-371A-18	Sequence 18, Appl
40		15	25.4	786431	4	US-09-751-389-3	Sequence 3, Appli
c 41		15	25.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
c 42		15	25.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c 43		15	25.4	1830121	4	US-10-329-960-1	Sequence 1, Appli
c 44		14	23.7	47	4	US-09-671-317-817	Sequence 817, App
45		14	23.7	133	5	PCT-US96-05611A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-864-357F-8
; Sequence 8, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 59
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-8

Query Match 100.0%; Score 59; DB 3; Length 59;
Best Local Similarity 100.0%; Pred. No. 5.5e-24;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59

RESULT 2
US-08-864-357F-11/c
; Sequence 11, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 59
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-11

Query Match 62.7%; Score 37; DB 3; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1
```

RESULT 3
US-08-961-527-171/c
; Sequence 171, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 171:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14736 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-171

Query Match 28.8%; Score 17; DB 4; Length 14736;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         42 TGAATCCATCATAAACT 58
              |||||||||||||||||
Db      13144 TGAATCCATCATAAACT 13128
```

RESULT 4
US-08-864-357F-10/c
; Sequence 10, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatc
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 60
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-10

Query Match 27.1%; Score 16; DB 3; Length 60;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         38 CGCGTGAATCCATCAT 53
              ||||||||||||||||
Db         60 CGCGTGAATCCATCAT 45
```

RESULT 5
US-09-252-991A-3153/c
; Sequence 3153, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3153
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3153

Query Match 27.1%; Score 16; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         24 CCTGCCGCAGAAACCG 39
              ||||||||||||||||
Db        465 CCTGCCGCAGAAACCG 450
```

RESULT 6
US-09-252-991A-810
; Sequence 810, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 810
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-810

Query Match 27.1%; Score 16; DB 4; Length 933;
Best Local Similarity 100.0%; Pred. No. 8.5;

```
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         18 TGACACCCTGCCGCAG 33
              ||||||||||||||||
Db        564 TGACACCCTGCCGCAG 579


RESULT 7
US-09-252-991A-869/c
; Sequence 869, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 869
;   LENGTH: 2142
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-869

  Query Match             27.1%;  Score 16;  DB 4;  Length 2142;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         18 TGACACCCTGCCGCAG 33
              ||||||||||||||||
Db       1639 TGACACCCTGCCGCAG 1624


RESULT 8
US-09-489-039A-1471
; Sequence 1471, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 1471
;   LENGTH: 309
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1471

  Query Match             25.4%;  Score 15;  DB 4;  Length 309;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCTGAAGAAACTGGT 17
              |||||||||||||||
Db        237 GCTGAAGAAACTGGT 251


RESULT 9
US-09-252-991A-3202
; Sequence 3202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 3202
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3202

  Query Match             25.4%;  Score 15;  DB 4;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         21 CACCCTGCCGCAGAA 35
              |||||||||||||||
Db          2 CACCCTGCCGCAGAA 16


RESULT 10
US-09-489-039A-1498/c
; Sequence 1498, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 1498
;   LENGTH: 369
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1498

  Query Match             25.4%;  Score 15;  DB 4;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCTGAAGAAACTGGT 17
              |||||||||||||||
Db        124 GCTGAAGAAACTGGT 110


RESULT 11
US-09-270-767-7800/c
; Sequence 7800, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7800
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-7800

  Query Match             25.4%;  Score 15;  DB 4;  Length 484;
```

```
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTG 15
              ||||||||||||||||
Db        366 CAGCTGAAGAAACTG 352


RESULT 12
US-09-270-767-23082/c
; Sequence 23082, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23082
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-23082

  Query Match             25.4%;  Score 15;  DB 4;  Length 484;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTG 15
              ||||||||||||||||
Db        366 CAGCTGAAGAAACTG 352


RESULT 13
US-09-270-767-1781/c
; Sequence 1781, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1781
;   LENGTH: 516
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-1781

  Query Match             25.4%;  Score 15;  DB 4;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTG 15
              ||||||||||||||||
Db        432 CAGCTGAAGAAACTG 418


RESULT 14
US-09-270-767-17063/c
; Sequence 17063, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17063
;   LENGTH: 516
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-17063

  Query Match             25.4%;  Score 15;  DB 4;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTG 15
              ||||||||||||||||
Db        432 CAGCTGAAGAAACTG 418


RESULT 15
US-09-489-039A-1476/c
; Sequence 1476, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 1476
;   LENGTH: 525
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1476

  Query Match             25.4%;  Score 15;  DB 4;  Length 525;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCTGAAGAAACTGGT 17
              ||||||||||||||||
Db        444 GCTGAAGAAACTGGT 430


Search completed: December 23, 2004, 01:33:42
Job time : 53.5204 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 23:36:53 ; Search time 827.068 Seconds
(without alignments)
397.214 Million cell updates/sec

Title: US-09-898-616A-3
Perfect score: 59
Sequence: 1 cagctgaagaaactggttga.....cgtgaatccatcataaactg 59

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4105333 seqs, 2784095677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8210666

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	59	100.0	59	9	US-09-861-688-8	Sequence 8, Appli
2	59	100.0	59	10	US-09-898-616A-3	Sequence 3, Appli
3	59	100.0	59	15	US-10-187-498A-3	Sequence 3, Appli
4	59	100.0	59	16	US-10-647-371-7	Sequence 7, Appli
c 5	37	62.7	59	9	US-09-861-688-11	Sequence 11, Appl
c 6	37	62.7	59	10	US-09-898-616A-6	Sequence 6, Appli
c 7	37	62.7	59	15	US-10-187-498A-6	Sequence 6, Appli
c 8	37	62.7	59	16	US-10-647-371-10	Sequence 10, Appl
9	19	32.2	757	17	US-10-437-963-89235	Sequence 89235, A
c 10	18	30.5	490	17	US-10-437-963-65306	Sequence 65306, A
c 11	18	30.5	1152	17	US-10-437-963-65308	Sequence 65308, A
c 12	18	30.5	2004	9	US-09-887-576-235	Sequence 235, App
13	18	30.5	2134	17	US-10-437-963-74313	Sequence 74313, A
c 14	17	28.8	195	9	US-09-974-300-3436	Sequence 3436, Ap
c 15	17	28.8	520	13	US-10-027-632-51350	Sequence 51350, A
c 16	17	28.8	520	15	US-10-027-632-51350	Sequence 51350, A
c 17	17	28.8	549	13	US-10-027-632-71225	Sequence 71225, A
c 18	17	28.8	549	13	US-10-027-632-312379	Sequence 312379,
c 19	17	28.8	549	15	US-10-027-632-71225	Sequence 71225, A
c 20	17	28.8	549	15	US-10-027-632-312379	Sequence 312379,
c 21	17	28.8	868	13	US-10-027-632-72285	Sequence 72285, A
c 22	17	28.8	868	15	US-10-027-632-72285	Sequence 72285, A
c 23	17	28.8	993	15	US-10-146-772-317	Sequence 317, App
c 24	17	28.8	993	16	US-10-241-742-317	Sequence 317, App
c 25	17	28.8	993	16	US-10-440-523-317	Sequence 317, App
c 26	17	28.8	993	16	US-10-440-503-317	Sequence 317, App
c 27	17	28.8	993	16	US-10-461-925-317	Sequence 317, App
c 28	17	28.8	1000	16	US-10-425-114-16587	Sequence 16587, A
c 29	17	28.8	1130	18	US-10-425-115-153378	Sequence 153378,
c 30	17	28.8	14736	8	US-08-961-527-171	Sequence 171, App
c 31	17	28.8	14736	16	US-10-158-844-171	Sequence 171, App
32	17	28.8	59475	17	US-10-322-696-166	Sequence 166, App
c 33	16	27.1	60	9	US-09-861-688-10	Sequence 10, Appl
c 34	16	27.1	60	10	US-09-898-616A-5	Sequence 5, Appli
c 35	16	27.1	60	15	US-10-187-498A-5	Sequence 5, Appli
c 36	16	27.1	60	16	US-10-647-371-9	Sequence 9, Appli
37	16	27.1	311	10	US-09-764-891-1420	Sequence 1420, Ap
38	16	27.1	373	16	US-10-424-599-11255	Sequence 11255, A
39	16	27.1	377	9	US-09-983-965-5306	Sequence 5306, Ap
40	16	27.1	455	18	US-10-425-115-115741	Sequence 115741,
c 41	16	27.1	547	16	US-10-424-599-35982	Sequence 35982, A
c 42	16	27.1	639	16	US-10-320-797-2026	Sequence 2026, Ap
c 43	16	27.1	659	18	US-10-653-047-581	Sequence 581, App
44	16	27.1	667	13	US-10-005-057A-34	Sequence 34, Appl
45	16	27.1	667	16	US-10-675-072A-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-09-861-688-8
; Sequence 8, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 59
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-8

Query Match 100.0%; Score 59; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.7e-23;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59

RESULT 2
US-09-898-616A-3

```
; Sequence 3, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-3

  Query Match           100.0%;  Score 59;  DB 10;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59


RESULT 3
US-10-187-498A-3
; Sequence 3, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-3

  Query Match           100.0%;  Score 59;  DB 15;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59


RESULT 4
US-10-647-371-7
; Sequence 7, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-7

  Query Match           100.0%;  Score 59;  DB 16;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-23;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAACCGCGTGAATCCATCATAAACTG 59


RESULT 5
US-09-861-688-11/c
; Sequence 11, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
;  TITLE OF INVENTION:  Inflammatory and
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE: 116142/2
;  CURRENT APPLICATION NUMBER: US/09/861,688
;  CURRENT FILING DATE:  2001-05-21
;  PRIOR APPLICATION NUMBER: 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-09-861-688-11

  Query Match           62.7%;  Score 37;  DB 9;  Length 59;
  Best Local Similarity  100.0%;  Pred. No. 5.4e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 6
US-09-898-616A-6/c
; Sequence 6, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-6

  Query Match             62.7%;  Score 37;  DB 10;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 7
US-10-187-498A-6/c
; Sequence 6, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-6

  Query Match             62.7%;  Score 37;  DB 15;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 8
US-10-647-371-10/c
; Sequence 10, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
;   LENGTH: 59
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-10

  Query Match             62.7%;  Score 37;  DB 16;  Length 59;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-11;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 37
              |||||||||||||||||||||||||||||||||||||
Db         37 CAGCTGAAGAAACTGGTTGACACCCTGCCGCAGAAAC 1


RESULT 9
US-10-437-963-89235
; Sequence 89235, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 89235
;   LENGTH: 757
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_88009C.1
US-10-437-963-89235

  Query Match             32.2%;  Score 19;  DB 17;  Length 757;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 AGCTGAAGAAACTGGTTGA 20
              |||||||||||||||||||
Db        207 AGCTGAAGAAACTGGTTGA 225


RESULT 10
US-10-437-963-65306/c
; Sequence 65306, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 65306
;   LENGTH: 490
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_66367C.1
US-10-437-963-65306

  Query Match            30.5%;  Score 18;  DB 17;  Length 490;
  Best Local Similarity  100.0%;  Pred. No. 3.8;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         13 CTGGTTGACACCCTGCCG 30
              ||||||||||||||||||
Db        335 CTGGTTGACACCCTGCCG 318


RESULT 11
US-10-437-963-65308/c
; Sequence 65308, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 65308
;   LENGTH: 1152
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_66369C.1
US-10-437-963-65308

  Query Match            30.5%;  Score 18;  DB 17;  Length 1152;
  Best Local Similarity  100.0%;  Pred. No. 4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         13 CTGGTTGACACCCTGCCG 30
              ||||||||||||||||||
Db        668 CTGGTTGACACCCTGCCG 651


RESULT 12
US-09-887-576-235/c
; Sequence 235, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 235
;   LENGTH: 2004
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-887-576-235

  Query Match            30.5%;  Score 18;  DB 9;  Length 2004;
  Best Local Similarity  100.0%;  Pred. No. 4.1;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         30 GCAGAAACCGCGTGAATC 47
              ||||||||||||||||||
Db        389 GCAGAAACCGCGTGAATC 372


RESULT 13
US-10-437-963-74313
; Sequence 74313, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 74313
;   LENGTH: 2134
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_74510C.1
US-10-437-963-74313

  Query Match            30.5%;  Score 18;  DB 17;  Length 2134;
  Best Local Similarity  100.0%;  Pred. No. 4.1;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCTGAAGAAACTGGTTGA 20
              ||||||||||||||||||
Db       1728 GCTGAAGAAACTGGTTGA 1745


RESULT 14
US-09-974-300-3436/c
; Sequence 3436, Application US/09974300
; Patent No. US20020146721A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3436
;   LENGTH: 195
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-3436

  Query Match             28.8%;  Score 17;  DB 9;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          21 CACCCTGCCGCAGAAAC 37
               |||||||||||||||||
Db         102 CACCCTGCCGCAGAAAC 86


RESULT 15
US-10-027-632-51350/c
; Sequence 51350, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 51350
;   LENGTH: 520
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-51350

  Query Match             28.8%;  Score 17;  DB 13;  Length 520;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           5 TGAAGAAACTGGTTGAC 21
               |||||||||||||||||
Db         106 TGAAGAAACTGGTTGAC 90


Search completed: December 23, 2004, 05:19:28
Job time : 829.068 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:17:58 ; Search time 523.357 Seconds
(without alignments)
3343.258 Million cell updates/sec

Title: US-09-898-616A-4
Perfect score: 37
Sequence: 1 atggagaagatcgctcagtctagcctgtgcaactaag 37

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		37	100.0	37	6	AR160918	AR160918 Sequence
2	c	37	100.0	60	6	AR160919	AR160919 Sequence
3	c	19	51.4	96850	2	AC102518_3	Continuation (4 of
4		19	51.4	110000	2	AC092450_5	Continuation (6 of
5	c	19	51.4	143233	2	AC102189	AC102189 Mus muscu
6	c	19	51.4	157842	9	AC092825	AC092825 Homo sapi
7		18	48.6	163934	9	AL691449	AL691449 Human DNA
8		18	48.6	167082	2	AP001903	AP001903 Homo sapi
9		18	48.6	168813	9	AP000555	AP000555 Homo sapi
10		18	48.6	169230	2	AC102275	AC102275 Mus muscu
11		18	48.6	172921	2	AP002454	AP002454 Homo sapi
12	c	18	48.6	177056	9	AC091843	AC091843 Homo sapi
13	c	18	48.6	180851	2	AC090210	AC090210 Homo sapi
14	c	18	48.6	181211	2	AP001904	AP001904 Homo sapi
15		18	48.6	186110	2	AC136510	AC136510 Pan trogl
16	c	18	48.6	204136	10	AL831776	AL831776 Mouse DNA
17		18	48.6	213552	9	AC010928	AC010928 Homo sapi
18		18	48.6	222545	2	AC109051	AC109051 Rattus no
19		18	48.6	228081	2	AC115237	AC115237 Rattus no
20	c	17	45.9	2997	9	AK127463	AK127463 Homo sapi
21	c	17	45.9	105856	9	AC015669	AC015669 Homo sapi
22		17	45.9	120817	2	AC136057	AC136057 Rattus no
23		17	45.9	149951	9	AC004970	AC004970 Homo sapi
24		17	45.9	156008	9	AL353093	AL353093 Human DNA
25	c	17	45.9	168602	2	AC124099	AC124099 Mus muscu
26	c	17	45.9	171112	9	AC016866	AC016866 Homo sapi
27		17	45.9	181936	2	AL360090	AL360090 Homo sapi
28		17	45.9	185472	2	AC128825	AC128825 Rattus no
29		17	45.9	201725	2	AL591706	AL591706 Homo sapi
30	c	17	45.9	206075	2	AC118511	AC118511 Rattus no
31	c	17	45.9	209380	10	AC133186	AC133186 Mus muscu
32	c	17	45.9	222162	10	AC109281	AC109281 Mus muscu
33	c	17	45.9	224226	2	AC129414	AC129414 Rattus no
34		17	45.9	235965	2	AC122659	AC122659 Rattus no
35		17	45.9	239924	2	AC111266	AC111266 Rattus no
36	c	17	45.9	240418	2	AC128419	AC128419 Rattus no
37		17	45.9	243598	2	AC142076	AC142076 Rattus no
38		17	45.9	243640	2	AC108571	AC108571 Rattus no
39		17	45.9	275059	2	AC120776	AC120776 Rattus no
40		16	43.2	666	11	BV047886	BV047886 S212P6112
41		16	43.2	1240	5	AF414086	AF414086 Xenopus l
42		16	43.2	2420	5	CID243835	AJ243835 Ctenophar
43		16	43.2	12282	1	AE010462	AE010462 Fusobacte
44		16	43.2	34667	5	BX571693	BX571693 Carp DNA
45	c	16	43.2	38490	9	AP001236	AP001236 Homo sapi

ALIGNMENTS

RESULT 1
AR160918
LOCUS AR160918 37 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 9 from patent US 6255281.
ACCESSION AR160918
VERSION AR160918.1 GI:16225987
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 37)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.
TITLE Use of recombinant human uteroglobin in treatment of inflammatory and fibrotic conditions
JOURNAL Patent: US 6255281-A 9 03-JUL-2001;
FEATURES Location/Qualifiers
source 1. .37
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 100.0%; Score 37; DB 6; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.7e-12;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37

RESULT 2
AR160919/c
LOCUS AR160919 60 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 10 from patent US 6255281.
ACCESSION AR160919
VERSION AR160919.1 GI:16225990
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 60)
AUTHORS Pilon,A.L., Mukherjee,A.B. and Zhang,Z.

```
  TITLE     Use of recombinant human uteroglobin in treatment of inflammatory
            and fibrotic conditions
  JOURNAL   Patent: US 6255281-A 10 03-JUL-2001;
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 37;  DB 6;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 3
AC102518_3/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC102518 Accession AC102518
     Fragment Name      Begin     End
     AC102518_0             1    110000
     AC102518_1        100001    210000
     AC102518_2        200001    310000
     AC102518_3        300001    396850
Continuation (4 of 4) of AC102518 from base 300001 (AC102518 Mus musculus chromosome 8 c

  Query Match             51.4%;  Score 19;  DB 2;  Length 96850;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          14 CTCAGTCTAGCCTGTGCAA 32
               ||||||||||||||||||
Db       22796 CTCAGTCTAGCCTGTGCAA 22778


RESULT 4
AC092450_5
WPCOMMENT
  Sequence split into 8 fragments  LOCUS AC092450 Accession AC092450
     Fragment Name      Begin     End
     AC092450_0             1    110000
     AC092450_1        100001    210000
     AC092450_2        200001    310000
     AC092450_3        300001    410000
     AC092450_4        400001    510000
     AC092450_5        500001    610000
     AC092450_6        600001    710000
     AC092450_7        700001    727300
Continuation (6 of 8) of AC092450 from base 500001 (AC092450 Homo sapiens chromosome 12

  Query Match             51.4%;  Score 19;  DB 2;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 0.75;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          14 CTCAGTCTAGCCTGTGCAA 32
               ||||||||||||||||||
Db       17216 CTCAGTCTAGCCTGTGCAA 17234


RESULT 5
AC102189/c
LOCUS       AC102189              143233 bp    DNA     linear   HTG 16-JUL-2003
DEFINITION  Mus musculus clone RP24-216J7, WORKING DRAFT SEQUENCE, 12 unordered
            pieces.
ACCESSION   AC102189
VERSION     AC102189.3  GI:32813512
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 143233)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-216J7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143233)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 143233)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 16, 2003 this sequence version replaced gi:22381210.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18187
                Center clone name: 216_J_7
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 139728 bases at least Q40
                Consensus quality: 141319 bases at least Q30
```

```
            Consensus quality: 141907 bases at least Q20
            Insert size: 133000; agarose-fp
            Insert size: 142133; sum-of-contigs
            Quality coverage: 10.7 in Q20 bases; agarose-fp
            Quality coverage: 10.0 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    10232: contig of 10232 bp in length
          *    10233    10332: gap of 100 bp
          *    10333    11435: contig of 1103 bp in length
          *    11436    11535: gap of 100 bp
          *    11536    12833: contig of 1298 bp in length
          *    12834    12933: gap of 100 bp
          *    12934    15356: contig of 2423 bp in length
          *    15357    15456: gap of 100 bp
          *    15457    17666: contig of 2210 bp in length
          *    17667    17766: gap of 100 bp
          *    17767    20013: contig of 2247 bp in length
          *    20014    20113: gap of 100 bp
          *    20114    23551: contig of 3438 bp in length
          *    23552    23651: gap of 100 bp
          *    23652    27815: contig of 4164 bp in length
          *    27816    27915: gap of 100 bp
          *    27916    36443: contig of 8528 bp in length
          *    36444    36543: gap of 100 bp
          *    36544    46767: contig of 10224 bp in length
          *    46768    46867: gap of 100 bp
          *    46868    66890: contig of 20023 bp in length
          *    66891    66990: gap of 100 bp
          *    66991   143233: contig of 76243 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143233
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-216J7"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .10232
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10333. .11435
                     /note="assembly_fragment"
     misc_feature    11536. .12833
                     /note="assembly_fragment"
     misc_feature    12934. .15356
                     /note="assembly_fragment"
     misc_feature    15457. .17666
                     /note="assembly_fragment"
     misc_feature    17767. .20013
                     /note="assembly_fragment"
     misc_feature    20114. .23551
                     /note="assembly_fragment"
     misc_feature    23652. .27815
                     /note="assembly_fragment"
     misc_feature    27916. .36443
                     /note="assembly_fragment"
     misc_feature    36544. .46767
                     /note="assembly_fragment"
     misc_feature    46868. .66890
                     /note="assembly_fragment"
     misc_feature    66991. .143233
                     /note="assembly_fragment"
ORIGIN

  Query Match          51.4%;  Score 19;  DB 2;  Length 143233;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         14 CTCAGTCTAGCCTGTGCAA 32
              ||||||||||||||||||
Db     105062 CTCAGTCTAGCCTGTGCAA 105044


RESULT 6
AC092825/c
LOCUS       AC092825              157842 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Homo sapiens 3 BAC RP11-171G7 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC092825
VERSION     AC092825.5  GI:21070529
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157842)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157842)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 157842)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
```

```
  JOURNAL    Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    4  (bases 1 to 157842)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (22-MAY-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On May 22, 2002 this sequence version replaced gi:20279217.
             INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
             gc-help@bcm.tmc.edu

             CLONE LENGTH: This sequence does not necessarily represent the
             entire insert of this clone. Overlapping regions of clones are only
             sequenced and submitted once, so the sequence for the remainder of
             the insert may be found in the record for the adjacent clones.
             Overlapping clones are noted at the beginning and end of the
             Features listing.

             ANNOTATION OF FEATURES:
                STSs are identified using ePCR (Genome Res. 7:541-550) searches
             of a local database that includes entries from dbSTS, GDB, and
             local mapping efforts.
                Repeats are identified using RepeatMasker (A. Smit and P. Green,
             unpublished.) for Human and Mouse sequences.
                Genes and Region of sequence similarity are identified by BLAST
             (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
             EST and cDNA sequences. Genes demonstrate at least two exons
             flanked by consensus splice sites that maintained sequence
             continuity across the splice junctions. Sequences that are not
             identical matches are annotated as similar.

             SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
             standard of double strand coverage with a minimum of 2 clones and 2
             reads with no ambiguities or 2 chemistries with a minimum of 2
             clones and 3 reads with no ambiguities. If the sequence quality for
             a region does not meet this standard, it will be indicated in the
             annotation as Low Coverage.

             QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
             standards - estimated error rate less than 1 per 10,000 bases.
             Reports of lowest quality individual bases and measures of base
             quality are listed below. Description of the metrics can be found
             at URL:
             http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

             QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .157842
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-171G7"
     repeat_region   163. .204
                     /rpt_family="T-rich"
     repeat_region   complement(206. .514)
                     /rpt_family="AluY"
     repeat_region   581. .811
                     /rpt_family="Charlie2"
     repeat_region   1005. .1307
                     /rpt_family="MLT2E"
     repeat_region   complement(1308. .1371)
                     /rpt_family="MER109"
     STS             2801. .3030
                     /standard_name="81156"
     repeat_region   2824. .2892
                     /rpt_family="(TCTA)n"
     repeat_region   2899. .2924
                     /rpt_family="(CA)n"
     repeat_region   complement(3582. .3979)
                     /rpt_family="MSTB"
     repeat_region   4295. .4334
                     /rpt_family="AT_rich"
     repeat_region   complement(7020. .7227)
                     /rpt_family="MER101"
     repeat_region   10026. .10204
                     /rpt_family="(TA)n"
     repeat_region   11515. .11546
                     /rpt_family="AT_rich"
     repeat_region   complement(13596. .13904)
                     /rpt_family="AluSx"
     repeat_region   complement(15095. .15165)
                     /rpt_family="Tigger1"
     repeat_region   15770. .15789
                     /rpt_family="(T)n"
     repeat_region   complement(15790. .17993)
                     /rpt_family="L1HS"
     STS             17161. .17338
                     /standard_name="61535"
     repeat_region   18421. .18480
                     /rpt_family="L2"
     repeat_region   complement(21210. .21359)
                     /rpt_family="L1"
     repeat_region   23843. .24385
                     /rpt_family="MER1A"
     repeat_region   complement(24743. .24873)
                     /rpt_family="L1M3a"
     repeat_region   complement(25825. .25997)
                     /rpt_family="L1M3a"
     repeat_region   26400. .26421
                     /rpt_family="AT_rich"
     repeat_region   complement(26586. .26814)
                     /rpt_family="L2"
     STS             27101. .27277
                     /standard_name="150162"
     repeat_region   complement(27621. .28060)
                     /rpt_family="L2"
     repeat_region   28243. .28408
                     /rpt_family="MIR"
     repeat_region   28577. .28697
                     /rpt_family="L2"
     repeat_region   30613. .30633
                     /rpt_family="AT_rich"
     repeat_region   complement(31014. .31089)
                     /rpt_family="MIR"
     repeat_region   complement(32739. .32905)
                     /rpt_family="MLT1I"
     repeat_region   33642. .33933
                     /rpt_family="AluSg"
     repeat_region   34226. .34535
                     /rpt_family="AluSx"
     repeat_region   35570. .35756
                     /rpt_family="MIR"
     repeat_region   36998. .37030
                     /rpt_family="(TA)n"
     repeat_region   37096. .37118
                     /rpt_family="(TTTTG)n"
     repeat_region   37244. .37285
                     /rpt_family="(TTTTG)n"
     repeat_region   41336. .41383
                     /rpt_family="MIR"
     repeat_region   41727. .41775
                     /rpt_family="AT_rich"
     repeat_region   42882. .43150
                     /rpt_family="AluJo"

  Query Match             51.4%;  Score 19;  DB 9;  Length 157842;
  Best Local Similarity   100.0%;  Pred. No. 0.73;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          14 CTCAGTCTAGCCTGTGCAA 32
               ||||||||||||||||||
Db      117040 CTCAGTCTAGCCTGTGCAA 117022
```

```
RESULT 7
AL691449
LOCUS       AL691449              163934 bp    DNA     linear   PRI 21-JAN-2003
DEFINITION  Human DNA sequence from clone RP11-558F24 on chromosome 1, complete
            sequence.
ACCESSION   AL691449
VERSION     AL691449.21  GI:27817297
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163934)
  AUTHORS   Matthews,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 21, 2003 this sequence version replaced gi:27803162.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-558F24 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .163934
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-558F24"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             48.6%;  Score 18;  DB 9;  Length 163934;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      44909 CAGTCTAGCCTGTGCAAC 44926


RESULT 8
AP001903
LOCUS       AP001903              167082 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-719K4 map 18q21, WORKING
            DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION   AP001903
VERSION     AP001903.2  GI:8117554
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167082)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 167,082 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 167082)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7649790.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-719K4
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 153234 bases at least Q40
                Consensus quality: 159520 bases at least Q30
                Consensus quality: 162786 bases at least Q20
                Insert size: 164582; sum-of-contigs
                Quality coverage: 5.73x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            26 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     22433 contig of  22433 bp in length
              22534     45017 contig of  22484 bp in length
              45118     59868 contig of  14751 bp in length
              59969     71793 contig of  11825 bp in length
              71894     81974 contig of  10081 bp in length
              82075     91440 contig of   9366 bp in length
              91541     98433 contig of   6893 bp in length
              98534    104713 contig of   6180 bp in length
             104814    108664 contig of   3851 bp in length
             108765    113626 contig of   4862 bp in length
             113727    120106 contig of   6380 bp in length
             120207    125002 contig of   4796 bp in length
             125103    129865 contig of   4763 bp in length
             129966    133326 contig of   3361 bp in length
             133427    138379 contig of   4953 bp in length
             138480    142887 contig of   4408 bp in length
             142988    146547 contig of   3560 bp in length
             146648    149530 contig of   2883 bp in length
             149631    152938 contig of   3308 bp in length
             153039    155111 contig of   2073 bp in length
             155212    157376 contig of   2165 bp in length
```

```
     157477    160277 contig of   2801 bp in length
     160378    162841 contig of   2464 bp in length
     162942    164597 contig of   1656 bp in length
     164698    166310 contig of   1613 bp in length
     166411    167082 contig of    672 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1    22433: contig of 22433 bp in length
*     22434    22533: gap of 100 bp
*     22534    45017: contig of 22484 bp in length
*     45018    45117: gap of 100 bp
*     45118    59868: contig of 14751 bp in length
*     59869    59968: gap of 100 bp
*     59969    71793: contig of 11825 bp in length
*     71794    71893: gap of 100 bp
*     71894    81974: contig of 10081 bp in length
*     81975    82074: gap of 100 bp
*     82075    91440: contig of 9366 bp in length
*     91441    91540: gap of 100 bp
*     91541    98433: contig of 6893 bp in length
*     98434    98533: gap of 100 bp
*     98534   104713: contig of 6180 bp in length
*    104714   104813: gap of 100 bp
*    104814   108664: contig of 3851 bp in length
*    108665   108764: gap of 100 bp
*    108765   113626: contig of 4862 bp in length
*    113627   113726: gap of 100 bp
*    113727   120106: contig of 6380 bp in length
*    120107   120206: gap of 100 bp
*    120207   125002: contig of 4796 bp in length
*    125003   125102: gap of 100 bp
*    125103   129865: contig of 4763 bp in length
*    129866   129965: gap of 100 bp
*    129966   133326: contig of 3361 bp in length
*    133327   133426: gap of 100 bp
*    133427   138379: contig of 4953 bp in length
*    138380   138479: gap of 100 bp
*    138480   142887: contig of 4408 bp in length
*    142888   142987: gap of 100 bp
*    142988   146547: contig of 3560 bp in length
*    146548   146647: gap of 100 bp
*    146648   149530: contig of 2883 bp in length
*    149531   149630: gap of 100 bp
*    149631   152938: contig of 3308 bp in length
*    152939   153038: gap of 100 bp
*    153039   155111: contig of 2073 bp in length
*    155112   155211: gap of 100 bp
*    155212   157376: contig of 2165 bp in length
*    157377   157476: gap of 100 bp
*    157477   160277: contig of 2801 bp in length
*    160278   160377: gap of 100 bp
*    160378   162841: contig of 2464 bp in length
*    162842   162941: gap of 100 bp
*    162942   164597: contig of 1656 bp in length
*    164598   164697: gap of 100 bp
*    164698   166310: contig of 1613 bp in length
*    166311   166410: gap of 100 bp
*    166411   167082: contig of 672 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167082
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-719K4"
     misc_feature    1. .22433
                     /note="assembly_fragment"
     misc_feature    22534. .45017
                     /note="assembly_fragment"
     misc_feature    45118. .59868
                     /note="assembly_fragment"
     misc_feature    59969. .71793
                     /note="assembly_fragment"
     misc_feature    71894. .81974
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    82075. .91440
                     /note="assembly_fragment"
     misc_feature    91541. .98433
                     /note="assembly_fragment"
     misc_feature    98534. .104713
                     /note="assembly_fragment"
     misc_feature    104814. .108664
                     /note="assembly_fragment"
     misc_feature    108765. .113626
                     /note="assembly_fragment"
     misc_feature    113727. .120106
                     /note="assembly_fragment"
     misc_feature    120207. .125002
                     /note="assembly_fragment"
     misc_feature    125103. .129865
                     /note="assembly_fragment"
     misc_feature    129966. .133326
                     /note="assembly_fragment"
     misc_feature    133427. .138379
                     /note="assembly_fragment"
     misc_feature    138480. .142887
                     /note="assembly_fragment"
     misc_feature    142988. .146547
                     /note="assembly_fragment"
     misc_feature    146648. .149530
                     /note="assembly_fragment"
     misc_feature    149631. .152938
                     /note="assembly_fragment"
     misc_feature    153039. .155111
                     /note="assembly_fragment"
     misc_feature    155212. .157376
                     /note="assembly_fragment"
     misc_feature    157477. .160277
                     /note="assembly_fragment"
     misc_feature    160378. .162841
                     /note="assembly_fragment"
     misc_feature    162942. .164597
                     /note="assembly_fragment"
     misc_feature    164698. .166310
                     /note="assembly_fragment"
     misc_feature    166411. .167082
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
ORIGIN

  Query Match          48.6%;  Score 18;  DB 2;  Length 167082;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      70348 CAGTCTAGCCTGTGCAAC 70365


RESULT 9
AP000555
LOCUS       AP000555              168813 bp    DNA     linear   PRI 01-OCT-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
            clone:KB1027C11.
ACCESSION   AP000555
VERSION     AP000555.1  GI:5931541
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168813)
AUTHORS Shimizu,N.
TITLE Human DNA sequence from clone KB1027C11 on chromosome 22q11.2
JOURNAL Published Only in DataBase (1999)
REFERENCE 2 (bases 1 to 168813)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1999) Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT This is a complete sequence of the insert of KB1027C11 clone. The proximal adjacent clone is KB666H9(Acc.#AP000554) with 586-bp overlapping. The distal adjacent clone is N109G12 (Acc.#D86995) with 24893-bp overlapping.
FEATURES Location/Qualifiers

source	1. .168813 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="22" /map="22q11.2" /clone="KB1027C11" /cell_line="FLEB14-14" /cell_type="pre-pro-B cell" /clone_lib="Keio BAC library"
repeat_region	3880. .4003 /evidence=not_experimental /rpt_family="MIR"
repeat_region	complement(6409. .6701) /evidence=not_experimental /rpt_family="AluYb8"
repeat_region	complement(6814. .6961) /evidence=not_experimental /rpt_family="FRAM"
repeat_region	7486. .7518 /evidence=not_experimental /rpt_family="AT_rich"
repeat_region	7899. .8175 /evidence=not_experimental /rpt_family="AluSx"
repeat_region	8454. .8739 /evidence=not_experimental /rpt_family="AluJo"
repeat_region	8742. .9041 /evidence=not_experimental /rpt_family="AluSc"
repeat_region	complement(9240. .9538) /evidence=not_experimental /rpt_family="AluSq"
repeat_region	complement(9542. .9833) /evidence=not_experimental /rpt_family="AluSx"
repeat_region	complement(9940. .10107) /evidence=not_experimental /rpt_family="AluJo"
repeat_region	complement(11679. .11989) /evidence=not_experimental /rpt_family="AluY"
repeat_region	12874. .13181 /evidence=not_experimental /rpt_family="AluY"
repeat_region	complement(13910. .14225) /evidence=not_experimental /rpt_family="AluJo"
repeat_region	complement(14232. .14525) /evidence=not_experimental /rpt_family="AluSg"
repeat_region	complement(14797. .14847) /evidence=not_experimental /rpt_family="Charlie1"
repeat_region	complement(14860. .15262) /evidence=not_experimental /rpt_family="Charlie1a"
repeat_region	15643. .15859 /evidence=not_experimental /rpt_family="L2"
repeat_region	complement(15899. .16209) /evidence=not_experimental /rpt_family="AluJb"
repeat_region	16465. .16486 /evidence=not_experimental /rpt_family="AT_rich"
repeat_region	16492. .16803 /evidence=not_experimental /rpt_family="AluY"
repeat_region	16838. .17049 /evidence=not_experimental /rpt_family="AluJo"
repeat_region	17247. .17332 /evidence=not_experimental /rpt_family="MLT1K"
repeat_region	17851. .17998 /evidence=not_experimental /rpt_family="L2"
repeat_region	complement(18114. .18296) /evidence=not_experimental /rpt_family="MIR"
repeat_region	complement(18503. .18635) /evidence=not_experimental /rpt_family="L1M4"
repeat_region	complement(18851. .18934) /evidence=not_experimental /rpt_family="L1P4"
repeat_region	19285. .19356 /evidence=not_experimental /rpt_family="L2"
repeat_region	19391. .19697 /evidence=not_experimental /rpt_family="AluSc"
repeat_region	19708. .19987 /evidence=not_experimental /rpt_family="AluSg"
repeat_region	19988. .20022 /evidence=not_experimental /rpt_family="(TAAA)n"
repeat_region	20035. .20363 /evidence=not_experimental /rpt_family="AluJo"
repeat_region	20558. .20711 /evidence=not_experimental /rpt_family="MLT1H"
repeat_region	20807. .21027 /evidence=not_experimental /rpt_family="AluJb"
repeat_region	21028. .21062 /evidence=not_experimental /rpt_family="AluY"
repeat_region	complement(21074. .21375) /evidence=not_experimental /rpt_family="AluJo"
repeat_region	complement(21472. .21611) /evidence=not_experimental /rpt_family="AluSg/x"
repeat_region	complement(21612. .21623) /evidence=not_experimental /rpt_family="AluY"
repeat_region	21624. .21643 /evidence=not_experimental /rpt_family="(TTTA)n"
repeat_region	complement(21644. .21913) /evidence=not_experimental /rpt_family="AluY"
repeat_region	21914. .22185

```
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   22190. .22505
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   22510. .22812
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   22872. .23141
                     /evidence=not_experimental
                     /rpt_family="AluSg1"
     repeat_region   complement(23261. .23563)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   23741. .24077
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   24095. .24405
                     /evidence=not_experimental
                     /rpt_family="AluY"
     repeat_region   24719. .24774
                     /evidence=not_experimental
                     /rpt_family="(CATATA)n"
     repeat_region   24915. .25015
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   complement(25077. .25132)
                     /evidence=not_experimental
                     /rpt_family="MLT1D"
     repeat_region   25191. .25245
                     /evidence=not_experimental
                     /rpt_family="MER66A"
     repeat_region   complement(25556. .25720)
                     /evidence=not_experimental
                     /rpt_family="L1M4"
     repeat_region   25717. .26058
                     /evidence=not_experimental
                     /rpt_family="L1ME1"
     repeat_region   26150. .26446
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   27486. .27783
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   27982. .28268
                     /evidence=not_experimental

  Query Match             48.6%;  Score 18;  DB 9;  Length 168813;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      73695 CAGTCTAGCCTGTGCAAC 73712


RESULT 10
AC102275
LOCUS       AC102275               169230 bp    DNA     linear   HTG 23-APR-2003
DEFINITION  Mus musculus clone RP24-290E7, WORKING DRAFT SEQUENCE, 40 unordered
            pieces.
ACCESSION   AC102275
VERSION     AC102275.2  GI:30017900
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 169230)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-290E7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169230)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 169230)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 17, 2003 this sequence version replaced gi:17061361.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18289
                Center clone name: 290_E_7
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 40 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   121664: contig of 121664 bp in length
            *   121665   121764: gap of 100 bp
            *   121765   122426: contig of 662 bp in length
```

```
*   122427   122526: gap of 100 bp
*   122527   123632: contig of 1106 bp in length
*   123633   123732: gap of 100 bp
*   123733   124478: contig of 746 bp in length
*   124479   124578: gap of 100 bp
*   124579   125233: contig of 655 bp in length
*   125234   125333: gap of 100 bp
*   125334   126416: contig of 1083 bp in length
*   126417   126516: gap of 100 bp
*   126517   127253: contig of 737 bp in length
*   127254   127353: gap of 100 bp
*   127354   128020: contig of 667 bp in length
*   128021   128120: gap of 100 bp
*   128121   128805: contig of 685 bp in length
*   128806   128905: gap of 100 bp
*   128906   129774: contig of 869 bp in length
*   129775   129874: gap of 100 bp
*   129875   130556: contig of 682 bp in length
*   130557   130656: gap of 100 bp
*   130657   131603: contig of 947 bp in length
*   131604   131703: gap of 100 bp
*   131704   132737: contig of 1034 bp in length
*   132738   132837: gap of 100 bp
*   132838   133498: contig of 661 bp in length
*   133499   133598: gap of 100 bp
*   133599   134340: contig of 742 bp in length
*   134341   134440: gap of 100 bp
*   134441   135339: contig of 899 bp in length
*   135340   135439: gap of 100 bp
*   135440   136302: contig of 863 bp in length
*   136303   136402: gap of 100 bp
*   136403   137116: contig of 714 bp in length
*   137117   137216: gap of 100 bp
*   137217   137842: contig of 626 bp in length
*   137843   137942: gap of 100 bp
*   137943   139264: contig of 1322 bp in length
*   139265   139364: gap of 100 bp
*   139365   140707: contig of 1343 bp in length
*   140708   140807: gap of 100 bp
*   140808   141956: contig of 1149 bp in length
*   141957   142056: gap of 100 bp
*   142057   142855: contig of 799 bp in length
*   142856   142955: gap of 100 bp
*   142956   143957: contig of 1002 bp in length
*   143958   144057: gap of 100 bp
*   144058   144736: contig of 679 bp in length
*   144737   144836: gap of 100 bp
*   144837   146268: contig of 1432 bp in length
*   146269   146368: gap of 100 bp
*   146369   147694: contig of 1326 bp in length
*   147695   147794: gap of 100 bp
*   147795   148673: contig of 879 bp in length
*   148674   148773: gap of 100 bp
*   148774   149716: contig of 943 bp in length
*   149717   149816: gap of 100 bp
*   149817   150915: contig of 1099 bp in length
*   150916   151015: gap of 100 bp
*   151016   152223: contig of 1208 bp in length
*   152224   152323: gap of 100 bp
*   152324   153902: contig of 1579 bp in length
*   153903   154002: gap of 100 bp
*   154003   155546: contig of 1544 bp in length
*   155547   155646: gap of 100 bp
*   155647   156697: contig of 1051 bp in length
*   156698   156797: gap of 100 bp
*   156798   158478: contig of 1681 bp in length
*   158479   158578: gap of 100 bp
*   158579   159753: contig of 1175 bp in length
*   159754   159853: gap of 100 bp
*   159854   161008: contig of 1155 bp in length
*   161009   161108: gap of 100 bp
*   161109   162235: contig of 1127 bp in length
*   162236   162335: gap of 100 bp
*   162336   163344: contig of 1009 bp in length
*   163345   163444: gap of 100 bp
*   163445   169230: contig of 5786 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169230
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-290E7"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .121664
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    121765. .122426
                     /note="assembly_fragment"
     misc_feature    122527. .123632
                     /note="assembly_fragment"
     misc_feature    123733. .124478
                     /note="assembly_fragment"
     misc_feature    124579. .125233
                     /note="assembly_fragment"
     misc_feature    125334. .126416
                     /note="assembly_fragment"
     misc_feature    126517. .127253
                     /note="assembly_fragment"
     misc_feature    127354. .128020
                     /note="assembly_fragment"
     misc_feature    128121. .128805
                     /note="assembly_fragment"
     misc_feature    128906. .129774
                     /note="assembly_fragment"
     misc_feature    129875. .130556
                     /note="assembly_fragment"
     misc_feature    130657. .131603
                     /note="assembly_fragment"
     misc_feature    131704. .132737
                     /note="assembly_fragment"
     misc_feature    132838. .133498

  Query Match             48.6%;  Score 18;  DB 2;  Length 169230;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          11 TCGCTCAGTCTAGCCTGT 28
               ||||||||||||||||||
Db       65094 TCGCTCAGTCTAGCCTGT 65111


RESULT 11
AP002454
LOCUS       AP002454               172921 bp    DNA     linear   HTG 06-JUN-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-819K4 map 18q21, WORKING
            DRAFT SEQUENCE, 29 unordered pieces.
ACCESSION   AP002454
VERSION     AP002454.1  GI:8307758
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172921)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 172,921 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 172921)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
```

```
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-819K4
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 152544 bases at least Q40
                Consensus quality: 162541 bases at least Q30
                Consensus quality: 167578 bases at least Q20
                Insert size: 170121; sum-of-contigs
                Quality coverage: 4.46x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            29 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     22518 contig of  22518 bp in length
                22619     36257 contig of  13639 bp in length
                36358     47612 contig of  11255 bp in length
                47713     57975 contig of  10263 bp in length
                58076     68186 contig of  10111 bp in length
                68287     74943 contig of   6657 bp in length
                75044     81573 contig of   6530 bp in length
                81674     88266 contig of   6593 bp in length
                88367     95348 contig of   6982 bp in length
                95449    101895 contig of   6447 bp in length
               101996    107285 contig of   5290 bp in length
               107386    113439 contig of   6054 bp in length
               113540    120229 contig of   6690 bp in length
               120330    126155 contig of   5826 bp in length
               126256    132029 contig of   5774 bp in length
               132130    136178 contig of   4049 bp in length
               136279    140578 contig of   4300 bp in length
               140679    144357 contig of   3679 bp in length
               144458    147973 contig of   3516 bp in length
               148074    151415 contig of   3342 bp in length
               151516    155725 contig of   4210 bp in length
               155826    158863 contig of   3038 bp in length
               158964    161765 contig of   2802 bp in length
               161866    164392 contig of   2527 bp in length
               164493    166109 contig of   1617 bp in length
               166210    168648 contig of   2439 bp in length
               168749    170747 contig of   1999 bp in length
               170848    172506 contig of   1659 bp in length
               172607    172921 contig of    315 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    22518: contig of 22518 bp in length
            *   22519    22618: gap of 100 bp
            *   22619    36257: contig of 13639 bp in length
            *   36258    36357: gap of 100 bp
            *   36358    47612: contig of 11255 bp in length
            *   47613    47712: gap of 100 bp
            *   47713    57975: contig of 10263 bp in length
            *   57976    58075: gap of 100 bp
            *   58076    68186: contig of 10111 bp in length
            *   68187    68286: gap of 100 bp
            *   68287    74943: contig of 6657 bp in length
            *   74944    75043: gap of 100 bp
            *   75044    81573: contig of 6530 bp in length
            *   81574    81673: gap of 100 bp
            *   81674    88266: contig of 6593 bp in length
            *   88267    88366: gap of 100 bp
            *   88367    95348: contig of 6982 bp in length
            *   95349    95448: gap of 100 bp
            *   95449   101895: contig of 6447 bp in length
            *  101896   101995: gap of 100 bp
            *  101996   107285: contig of 5290 bp in length
            *  107286   107385: gap of 100 bp
            *  107386   113439: contig of 6054 bp in length
            *  113440   113539: gap of 100 bp
            *  113540   120229: contig of 6690 bp in length
            *  120230   120329: gap of 100 bp
            *  120330   126155: contig of 5826 bp in length
            *  126156   126255: gap of 100 bp
            *  126256   132029: contig of 5774 bp in length
            *  132030   132129: gap of 100 bp
            *  132130   136178: contig of 4049 bp in length
            *  136179   136278: gap of 100 bp
            *  136279   140578: contig of 4300 bp in length
            *  140579   140678: gap of 100 bp
            *  140679   144357: contig of 3679 bp in length
            *  144358   144457: gap of 100 bp
            *  144458   147973: contig of 3516 bp in length
            *  147974   148073: gap of 100 bp
            *  148074   151415: contig of 3342 bp in length
            *  151416   151515: gap of 100 bp
            *  151516   155725: contig of 4210 bp in length
            *  155726   155825: gap of 100 bp
            *  155826   158863: contig of 3038 bp in length
            *  158864   158963: gap of 100 bp
            *  158964   161765: contig of 2802 bp in length
            *  161766   161865: gap of 100 bp
            *  161866   164392: contig of 2527 bp in length
            *  164393   164492: gap of 100 bp
            *  164493   166109: contig of 1617 bp in length
            *  166110   166209: gap of 100 bp
            *  166210   168648: contig of 2439 bp in length
            *  168649   168748: gap of 100 bp
            *  168749   170747: contig of 1999 bp in length
            *  170748   170847: gap of 100 bp
            *  170848   172506: contig of 1659 bp in length
            *  172507   172606: gap of 100 bp
            *  172607   172921: contig of 315 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172921
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-819K4"
     misc_feature    1. .22518
                     /note="assembly_fragment"
     misc_feature    22619. .36257
                     /note="assembly_fragment"
     misc_feature    36358. .47612
                     /note="assembly_fragment"
     misc_feature    47713. .57975
                     /note="assembly_fragment"
     misc_feature    58076. .68186
                     /note="assembly_fragment"
     misc_feature    68287. .74943
                     /note="assembly_fragment"
     misc_feature    75044. .81573
                     /note="assembly_fragment"
     misc_feature    81674. .88266
                     /note="assembly_fragment"
```

```
     misc_feature    88367. .95348
                     /note="assembly_fragment"
     misc_feature    95449. .101895
                     /note="assembly_fragment"
     misc_feature    101996. .107285
                     /note="assembly_fragment"
     misc_feature    107386. .113439
                     /note="assembly_fragment"
     misc_feature    113540. .120229
                     /note="assembly_fragment"
     misc_feature    120330. .126155
                     /note="assembly_fragment"
     misc_feature    126256. .132029
                     /note="assembly_fragment"
     misc_feature    132130. .136178
                     /note="assembly_fragment"
     misc_feature    136279. .140578
                     /note="assembly_fragment"
     misc_feature    140679. .144357
                     /note="assembly_fragment"
     misc_feature    144458. .147973
                     /note="assembly_fragment"
     misc_feature    148074. .151415
                     /note="assembly_fragment"
     misc_feature    151516. .155725
                     /note="assembly_fragment"
     misc_feature    155826. .158863
                     /note="assembly_fragment"
     misc_feature    158964. .161765
                     /note="assembly_fragment"
     misc_feature    161866. .164392
                     /note="assembly_fragment"

  Query Match             48.6%;  Score 18;  DB 2;  Length 172921;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      34753 CAGTCTAGCCTGTGCAAC 34770


RESULT 12
AC091843/c
LOCUS       AC091843               177056 bp    DNA     linear   PRI 27-FEB-2002
DEFINITION  Homo sapiens chromosome 5 clone CTD-2185O23, complete sequence.
ACCESSION   AC091843
VERSION     AC091843.3  GI:18958638
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177056)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177056)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 177056)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:15290412.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .177056
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2185O23"
ORIGIN

  Query Match             48.6%;  Score 18;  DB 9;  Length 177056;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db     146662 CAGTCTAGCCTGTGCAAC 146645


RESULT 13
AC090210/c
LOCUS       AC090210               180851 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens chromosome 18 clone RP11-720N18 map 18.
ACCESSION   AC090210
VERSION     AC090210.4  GI:15290848
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180851)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-720N18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180851)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 180851)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
```

```
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 26, 2001 this sequence version replaced gi:14192981.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12233
                Center clone name: 720_N_18
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   180851: contig of 180851 bp in length.
FEATURES            Location/Qualifiers
     source         1. .180851
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="18"
                    /map="18"
                    /clone="RP11-720N18"
                    /clone_lib="RPCI-11 Human Male BAC"
ORIGIN

  Query Match             48.6%;  Score 18;  DB 2;  Length 180851;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      33216 CAGTCTAGCCTGTGCAAC 33199


RESULT 14
AP001904/c
LOCUS       AP001904               181211 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-720N18 map 18q21, WORKING
            DRAFT SEQUENCE, 15 unordered pieces.
ACCESSION   AP001904
VERSION     AP001904.2  GI:8117555
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181211)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 181,211 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 181211)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7649791.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-720N18
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 171968 bases at least Q40
                Consensus quality: 176258 bases at least Q30
                Consensus quality: 177945 bases at least Q20
                Insert size: 179811; sum-of-contigs
                Quality coverage: 5.94x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            15 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved.
                  1     33300 contig of  33300 bp in length
              33401     54981 contig of  21581 bp in length
              55082     73468 contig of  18387 bp in length
              73569     94430 contig of  20862 bp in length
              94531    113614 contig of  19084 bp in length
             113715    123984 contig of  10270 bp in length
             124085    137815 contig of  13731 bp in length
             137916    147429 contig of   9514 bp in length
             147530    154898 contig of   7369 bp in length
             154999    162210 contig of   7212 bp in length
             162311    168580 contig of   6270 bp in length
             168681    171531 contig of   2851 bp in length
             171632    175227 contig of   3596 bp in length
             175328    178652 contig of   3325 bp in length
             178753    181211 contig of   2459 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     33300: contig of 33300 bp in length
            *   33301     33400: gap of 100 bp
            *   33401     54981: contig of 21581 bp in length
            *   54982     55081: gap of 100 bp
            *   55082     73468: contig of 18387 bp in length
            *   73469     73568: gap of 100 bp
            *   73569     94430: contig of 20862 bp in length
            *   94431     94530: gap of 100 bp
            *   94531    113614: contig of 19084 bp in length
            *  113615    113714: gap of 100 bp
            *  113715    123984: contig of 10270 bp in length
            *  123985    124084: gap of 100 bp
            *  124085    137815: contig of 13731 bp in length
            *  137816    137915: gap of 100 bp
```

```
            *   137916   147429: contig of 9514 bp in length
            *   147430   147529: gap of 100 bp
            *   147530   154898: contig of 7369 bp in length
            *   154899   154998: gap of 100 bp
            *   154999   162210: contig of 7212 bp in length
            *   162211   162310: gap of 100 bp
            *   162311   168580: contig of 6270 bp in length
            *   168581   168680: gap of 100 bp
            *   168681   171531: contig of 2851 bp in length
            *   171532   171631: gap of 100 bp
            *   171632   175227: contig of 3596 bp in length
            *   175228   175327: gap of 100 bp
            *   175328   178652: contig of 3325 bp in length
            *   178653   178752: gap of 100 bp
            *   178753   181211: contig of 2459 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181211
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-720N18"
     misc_feature    1. .33300
                     /note="assembly_fragment"
     misc_feature    33401. .54981
                     /note="assembly_fragment"
     misc_feature    55082. .73468
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    73569. .94430
                     /note="assembly_fragment"
     misc_feature    94531. .113614
                     /note="assembly_fragment"
     misc_feature    113715. .123984
                     /note="assembly_fragment"
     misc_feature    124085. .137815
                     /note="assembly_fragment"
     misc_feature    137916. .147429
                     /note="assembly_fragment"
     misc_feature    147530. .154898
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    154999. .162210
                     /note="assembly_fragment"
     misc_feature    162311. .168580
                     /note="assembly_fragment"
     misc_feature    168681. .171531
                     /note="assembly_fragment"
     misc_feature    171632. .175227
                     /note="assembly_fragment"
     misc_feature    175328. .178652
                     /note="assembly_fragment"
     misc_feature    178753. .181211
                     /note="assembly_fragment"
ORIGIN

  Query Match             48.6%;  Score 18;  DB 2;  Length 181211;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db     180151 CAGTCTAGCCTGTGCAAC 180134


RESULT 15
AC136510
LOCUS       AC136510              186110 bp    DNA     linear   HTG 18-MAR-2004
DEFINITION  Pan troglodytes clone rp43-27h3, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC136510
VERSION     AC136510.17  GI:45544719
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 186110)
  AUTHORS   Lau,C.C.Y. and Roe,B.A.
  TITLE     Pan troglodytes BAC Clone rp43-27h3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186110)
  AUTHORS   Lau,C.C.Y. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 186110)
  AUTHORS   Lau,C.C.Y. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Mar 18, 2004 this sequence version replaced gi:45476613.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    76466: contig of 76466 bp in length
            *    76467    76566: gap of unknown length
            *    76567    82125: contig of 5559 bp in length
            *    82126    82225: gap of unknown length
            *    82226    92187: contig of 9962 bp in length
            *    92188    92287: gap of unknown length
            *    92288    94597: contig of 2310 bp in length
            *    94598    94697: gap of unknown length
            *    94698   112016: contig of 17319 bp in length
            *   112017   112116: gap of unknown length
            *   112117   186110: contig of 73994 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186110
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="rp43-27h3"
                     /clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
ORIGIN

  Query Match             48.6%;  Score 18;  DB 2;  Length 186110;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTGCAAC 33
              ||||||||||||||||||
Db      36383 CAGTCTAGCCTGTGCAAC 36400


Search completed: December 22, 2004, 23:36:46
Job time : 531.357 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 19:14:28 ; Search time 130.588 Seconds
(without alignments)
1487.336 Million cell updates/sec

Title: US-09-898-616A-4
Perfect score: 37
Sequence: 1 atggagaagatcgctcagtctagcctgtgcaactaag 37

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	37	100.0	37	9	ABZ58373	Abz58373 Human ute
	2	37	100.0	37	12	ADL27629	Adl27629 Recombina
c	3	37	100.0	60	9	ABZ58374	Abz58374 Human ute
c	4	37	100.0	60	12	ADL27630	Adl27630 Recombina
c	5	16	43.2	698	4	AAK69658	Aak69658 Human imm
c	6	16	43.2	699	4	AAK69659	Aak69659 Human imm
c	7	16	43.2	699	4	AAK69660	Aak69660 Human imm
	8	15	40.5	477	9	ACH14100	Ach14100 Human adu
	9	15	40.5	621	4	AAI92367	Aai92367 Human pol
	10	15	40.5	675	10	ADD35086	Add35086 Mouse mit
c	11	15	40.5	1043	2	AAZ24882	Aaz24882 Human sec
	12	15	40.5	1089	2	AAX37460	Aax37460 Human sec
	13	15	40.5	1601	3	AAA28524	Aaa28524 Human opi
c	14	15	40.5	1756	10	ADD44906	Add44906 Rat gene
	15	15	40.5	2289	3	AAA28526	Aaa28526 Human opi
	16	15	40.5	2290	3	AAA28522	Aaa28522 Human opi
	17	15	40.5	2348	3	AAA28525	Aaa28525 Human opi
	18	15	40.5	2408	3	AAA28523	Aaa28523 Human opi
	19	15	40.5	2409	4	AAH17441	Aah17441 Human cDN
	20	15	40.5	2423	10	ADD14783	Add14783 Human src
c	21	15	40.5	3460	10	ADE56935	Ade56935 Rat gene
	22	15	40.5	3715	8	ADA15021	Ada15021 Murine an
	23	15	40.5	4065	4	AAD16913	Aad16913 Human pho
c	24	15	40.5	8003	6	ABA99404	Aba99404 B. mori t
	25	15	40.5	13660	4	AAI99126	Aai99126 Human exc
	26	15	40.5	13660	4	AAK81002	Aak81002 Human imm
	27	15	40.5	13660	4	AAK67907	Aak67907 Human imm
	28	15	40.5	13660	5	AAI63476	Aai63476 Human kid
c	29	15	40.5	16122	4	AAS22908	Aas22908 DNA encod
c	30	15	40.5	16341	4	AAS23002	Aas23002 DNA encod
c	31	15	40.5	28001	12	ADI36729	Adi36729 Genomic D
c	32	15	40.5	28001	12	ADI36730	Adi36730 Genomic D
c	33	15	40.5	28001	12	ADM93169	Adm93169 Human KOX
c	34	15	40.5	28001	12	ADM93170	Adm93170 Human KOX
c	35	15	40.5	78025	8	ABQ77404	Abq77404 Human SEL
c	36	15	40.5	153170	12	ADQ17382	Adq17382 Human sof
	37	15	40.5	183610	8	ACF62736	Acf62736 Cancer ba
	38	15	40.5	183610	8	ADB20851	Adb20851 MRP1 base
	39	15	40.5	183610	10	ADB87940	Adb87940 Human UGT
	40	15	40.5	183610	10	ADB96923	Adb96923 Human MDR
	41	15	40.5	183610	10	ADB92114	Adb92114 Human MDR
	42	15	40.5	186591	8	ACF62750	Acf62750 Cancer ba
	43	15	40.5	186591	8	ADB20869	Adb20869 MRP1 base
	44	15	40.5	186591	10	ADB87958	Adb87958 Human UGT
	45	15	40.5	186591	10	ADB96941	Adb96941 Human MDR

ALIGNMENTS

RESULT 1
ABZ58373
ID ABZ58373 standard; DNA; 37 BP.
XX
AC ABZ58373;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 4.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Claim 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 4, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the
CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and

CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 37 BP; 11 A; 8 C; 10 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 37; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 4.8e-11;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37

RESULT 2
ADL27629
ID ADL27629 standard; DNA; 37 BP.
XX
AC ADL27629;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, coding oligonucleotide #4.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Claim 1; SEQ ID NO 4; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a coding strand oligonucleotide used to
CC construct the synthetic rhUG gene.
XX
SQ Sequence 37 BP; 11 A; 8 C; 10 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 37; DB 12; Length 37;
Best Local Similarity 100.0%; Pred. No. 4.8e-11;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37

RESULT 3
ABZ58374/c
ID ABZ58374 standard; DNA; 60 BP.
XX
AC ABZ58374;
XX
DT 28-APR-2003 (first entry)
XX
DE Human uteroglobin synthetic gene oligonucleotide 5.
XX
KW Human; uteroglobin; respiratory distress; antiinflammatory; antifibrotic;
KW antiinflammatory; antiasthmatic; nephrotropic; antirheumatic;
KW antiarthritic; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003003979-A2.
XX
PD 16-JAN-2003.
XX
PF 02-JUL-2002; 2002WO-US020836.
XX
PR 02-JUL-2001; 2001US-00898616.
XX
PA (CLAR-) CLARAGEN INC.
XX
PI Pilon AL, Welch RE;
XX
DR WPI; 2003-221527/21.
XX
PT Bacterial expression system for producing recombinant human uteroglobin
PT for treating inflammatory and fibrotic conditions, comprises a synthetic
PT gene which codes for human uteroglobin.
XX
PS Example 1; Page 33; 127pp; English.
XX
CC The present sequence is that of oligonucleotide 5, which was used in the
CC construction of a synthetic gene for the production of human uteroglobin
CC (hUG) in bacteria. Oligonucleotides 1-4 (see ABZ58370-73) were used to
CC assemble the coding strand and oligonucleotides 5-8 (see ABZ58374-77) the

CC complementary strand. The gene was assembled by annealing and ligation of
CC the oligonucleotides. Because mature native hUG has glutamic acid at its
CC N-terminus, an initiator methionine was added to the N-terminus, and
CC codon usage was optimised for expression in bacteria. In an example from
CC the invention, the synthetic gene was cloned into plasmid pCG12 (see
CC ABZ58378) and recombinant hUG (see ABP72259) was produced in Escherichia
CC coli strain CG12. The invention relates generally to the production of
CC recombinant hUG by bacterial expression, protein purification and scaled-
CC up production according to current good manufacturing practices. The
CC recombinant hUG is useful for the treatment of inflammatory and fibrotic
CC conditions, such as neonatal respiratory distress syndrome and
CC bronchopulmonary dysplasia. It may also be used to treat conditions
CC associated with elevated phospholipase A2 levels such as pancreatitis,
CC acute renal failure, rheumatoid arthritis and asthma
XX
SQ Sequence 60 BP; 13 A; 14 C; 15 G; 18 T; 0 U; 0 Other;

Query Match 100.0%; Score 37; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.8e-11;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1

RESULT 4
ADL27630/c
ID ADL27630 standard; DNA; 60 BP.
XX
AC ADL27630;
XX
DT 20-MAY-2004 (first entry)
XX
DE Recombinant human uteroglobin, rhUG, non-coding oligonucleotide #1.
XX
KW Human; ss; recombinant human uteroglobin; rhUG;
KW bacterial expression system; rhUG master cell bank;
KW rhUG research seed bank; anti-inflammatory; secretory phospholipase A 2;
KW fibronectin; respiratory distress; inflammation; fibrotic disease.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003207795-A1.
XX
PD 06-NOV-2003.
XX
PF 02-JUL-2002; 2002US-00187498.
XX
PR 28-MAY-1997; 97US-00864357.
PR 02-JUL-2001; 2001US-00898616.
XX
PA (PILO/) PILON A L.
PA (WELC/) WELCH R W.
XX
PI Pilon AL, Welch RW;
XX
DR WPI; 2004-051527/05.
XX
PT Bacterial expression system for production of recombinant human
PT uteroglobin comprising synthetic gene or human cDNA sequence which codes
PT for human uteroglobin.
XX
PS Example 1; SEQ ID NO 5; 64pp; English.
XX
CC The invention relates to a bacterial expression system for the production
CC of recombinant human uteroglobin (rhUG), comprising a synthetic gene or
CC human cDNA sequence which codes for human UG, constructed from the
CC oligonucleotides appearing as ADL27626-ADL27629, and which further
CC comprises Met-Ala-Ala at the N-terminus of the sequence. Also included
CC are producing an rhUG master cell bank (comprising inoculating a suitable
CC incubating broth with an aliquot portion of a rhUG research seed bank to
CC form a bacterial culture, incubating the bacterial culture, adding a
CC cryopreservative to the bacterial culture to form a cryopreserved
CC solution, transferring a portion of the cryopreserved solution to a
CC cryovial and storing the cryovial at a temperature below -60 degrees C),
CC expressing rhUG (comprising providing a production seed cell bank culture
CC comprising an expression vector capable of expressing rhUG, inoculating a
CC broth medium with the production seed cell bank culture to form an
CC inoculum, incubating the bacterial culture formed in step (b),
CC inoculating a large scale fermenter with the inoculum formed from the
CC step (c) to form a fermentation culture, incubating the fermentation
CC culture within the large scale fermenter, adding an induction agent to
CC the fermentation culture to induce the expression of rhUG and harvesting
CC the above fermentation culture), purifying rhUG, determining the potency
CC of rhUG in a sample, measuring in vitro anti-inflammatory effect arising
CC from inhibition or blocking of secretory phospholipase A 2 enzymes by
CC rhUG, measuring in vitro binding of rhUG to fibronectin, determining the
CC purity of rhUG, and a pharmaceutical composition comprising a purified
CC rhUG and a carrier or diluent. The bacterial expression system is useful
CC for producing a rhUG research seed bank or a pharmaceutical grade rhUG
CC drug substance. rhUG is safe to administer to a patient in respiratory
CC distress. The rhUG is useful for treating inflammation and fibrotic
CC diseases. The present sequence is a non-coding strand oligonucleotide
CC used to construct the synthetic rhUG gene.
XX
SQ Sequence 60 BP; 13 A; 14 C; 15 G; 18 T; 0 U; 0 Other;

Query Match 100.0%; Score 37; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.8e-11;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1

RESULT 5
AAK69658/c
ID AAK69658 standard; DNA; 698 BP.
XX
AC AAK69658;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24470.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 24470; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic

```
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 698 BP; 270 A; 116 C; 107 G; 205 T; 0 U; 0 Other;

  Query Match             43.2%;  Score 16;  DB 4;  Length 698;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         17 AGTCTAGCCTGTGCAA 32
              ||||||||||||||||
Db        664 AGTCTAGCCTGTGCAA 649


RESULT 6
AAK69659/c
ID   AAK69659 standard; DNA; 699 BP.
XX
AC   AAK69659;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24471.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
```

```
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 24471; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 699 BP; 270 A; 116 C; 107 G; 206 T; 0 U; 0 Other;

  Query Match             43.2%;  Score 16;  DB 4;  Length 699;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         17 AGTCTAGCCTGTGCAA 32
              ||||||||||||||||
Db        665 AGTCTAGCCTGTGCAA 650


RESULT 7
AAK69660/c
ID   AAK69660 standard; DNA; 699 BP.
XX
AC   AAK69660;
XX
DT   06-NOV-2001  (first entry)

XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24472.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
```

PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 24472; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 699 BP; 270 A; 115 C; 107 G; 207 T; 0 U; 0 Other;

Query Match 43.2%; Score 16; DB 4; Length 699;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 AGTCTAGCCTGTGCAA 32
||||||||||||||||
Db 665 AGTCTAGCCTGTGCAA 650

RESULT 8
ACH14100
ID ACH14100 standard; cDNA; 477 BP.
XX
AC ACH14100;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult brain cDNA #1312.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.

```
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 1312; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 477 BP; 101 A; 116 C; 194 G; 52 T; 0 U; 14 Other;

  Query Match             40.5%;  Score 15;  DB 9;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TGGAGAAGATCGCTC 16
              |||||||||||||||
Db        305 TGGAGAAGATCGCTC 319


RESULT 9
AAI92367
ID   AAI92367 standard; cDNA; 621 BP.
XX
AC   AAI92367;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 12427.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO12436.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 1; SEQ ID NO 12427; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 621 BP; 152 A; 162 C; 137 G; 168 T; 0 U; 2 Other;

  Query Match             40.5%;  Score 15;  DB 4;  Length 621;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAAC 33
              |||||||||||||||
Db        113 TCTAGCCTGTGCAAC 127


RESULT 10
ADD35086
ID   ADD35086 standard; DNA; 675 BP.
XX
AC   ADD35086;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Mouse mitochondrial DNA sequence SEQ ID NO:2866.
XX
KW   ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;
KW   mitochondrial disease; oxidative phosphorylation dysfunction;
KW   oxidative stress; apoptosis; aging.
XX
OS   Mus musculus.
XX
PN   WO2003020220-A2.
XX
PD   13-MAR-2003.
XX
PF   30-AUG-2002; 2002WO-US027886.
XX
PR   30-AUG-2001; 2001US-0316323P.
PR   31-AUG-2001; 2001CA-02356540.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wallace DC,  Levy S,  Kerstann K,  Procaccio V;
XX
DR   WPI; 2003-300821/29.
XX
PT   Array containing probes for genes involved in mitochondrial biology,
PT   useful for determining mitochondrial biology gene expression profiles for
PT   use in diagnosing pathologies and identifying biochemical pathways.
XX
PS   Claim 2; SEQ ID NO 2866; 201pp; English.
XX
CC   The invention relates to a novel array comprising at least two isolated
CC   nucleotide molecules, each molecule having a sequence capable of uniquely
CC   hybridising to a nucleic acid molecule which is an expression product of
CC   a gene involved in mitochondrial biology. The array comprises two or more
CC   isolated nucleic acid molecules or spots, each molecule having a sequence
```

CC chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC of the invention is useful for determining an expression profile of a
CC mouse or human sample containing nucleic acid, by contacting the array
CC with the sample under conditions allowing selective hybridisation, and
CC measuring hybridisation of nucleic acid in the sample to the array to
CC produce an expression profile. The array is also useful for determining
CC an expression profile of a first labelled sample containing nucleic acid
CC relative to a second, differently labelled sample containing nucleic
CC acid. The second sample is a reference or a standard. An array is useful
CC for determining an expression profile diagnostic of an energy-metabolism-
CC related physiological condition. An array of the invention is useful for
CC determining mitochondrial biology gene expression profiles of organisms,
CC such as human, mice and closely related species, tissue and organs of
CC such organisms, which are useful for determining expression profiles
CC diagnostic of energy metabolism-related physiological conditions,
CC diagnosing such physiological conditions, identifying biochemical
CC pathways, genes, and mutations involved in such physiological conditions,
CC identifying therapeutic agents useful for preventing and/or treating such
CC physiological conditions, evaluating and/or monitoring the efficacy of
CC such therapies, and creating and identifying animal models of human
CC energy metabolism-related physiological conditions. An array is also
CC useful for defining expression signatures or profiles for mitochondrial
CC diseases, as well as distinguishing clinical disorders that result from
CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC apoptosis and aging. An array of the invention contains probes of genes
CC not previously recognised to participate in mitochondrial biology. The
CC sequences shown in ADD33224-ADD35260 represent murine mitochondrial DNA
CC clones used to make the probes of the invention. Some sequences are not
CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC 1906, 2408 and 2643.
XX
SQ Sequence 675 BP; 173 A; 142 C; 129 G; 231 T; 0 U; 0 Other;

Query Match 40.5%; Score 15; DB 10; Length 675;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 TCAGTCTAGCCTGTG 29
|||||||||||||||
Db 301 TCAGTCTAGCCTGTG 315

RESULT 11
AAZ24882/c
ID AAZ24882 standard; DNA; 1043 BP.
XX
AC AAZ24882;
XX
DT 02-DEC-1999 (first entry)
XX
DE Human secreted protein gene 72 clone HE6GA29.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9947540-A1.
XX
PD 23-SEP-1999.
XX
PF 18-MAR-1999; 99WO-US005804.
XX
PR 19-MAR-1998; 98US-0078563P.
PR 19-MAR-1998; 98US-0078566P.
PR 19-MAR-1998; 98US-0078573P.
PR 19-MAR-1998; 98US-0078574P.
PR 19-MAR-1998; 98US-0078576P.
PR 19-MAR-1998; 98US-0078577P.
PR 19-MAR-1998; 98US-0078578P.
PR 19-MAR-1998; 98US-0078579P.
PR 19-MAR-1998; 98US-0078581P.
PR 01-APR-1998; 98US-0080312P.
PR 01-APR-1998; 98US-0080313P.
PR 01-APR-1998; 98US-0080314P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;
PI Shi Y, Moore PA;
XX
DR WPI; 1999-562050/47.
DR P-PSDB; AAY41379.
XX
PT New isolated human genes, useful for diagnosis and treatment of e.g.
PT cancers, neurological disorders, immune diseases, inflammation or blood
PT disorders.
XX
PS Claim 1; Page 341; 484pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. AAZ24802) for increasing the stability of the fused protein as
CC compared to the human protein only. The invention relates to 95 novel
CC genes and their fragments (nucleic acid sequences: AAZ24811-Z24907; amino
CC acid sequences AAY41308-Y41404) which are useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
CC pathological conditions can be diagnosed by determining the amount of the
CC new polypeptides in a sample or by determining the presence of mutations
CC in the new polynucleotides. Specific uses are described for each of the
CC 95 polynucleotides, based on which tissues they are most highly expressed
CC in (see AAZ24811 for described uses)
XX
SQ Sequence 1043 BP; 281 A; 236 C; 291 G; 235 T; 0 U; 0 Other;

Query Match 40.5%; Score 15; DB 2; Length 1043;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 TCTAGCCTGTGCAAC 33
|||||||||||||||
Db 597 TCTAGCCTGTGCAAC 583

RESULT 12
AAX37460
ID AAX37460 standard; cDNA; 1089 BP.
XX
AC AAX37460;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 10.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN WO9918208-A1.
XX

PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US020775.
XX
PR 02-OCT-1997; 97US-0060833P.
PR 02-OCT-1997; 97US-0060836P.
PR 02-OCT-1997; 97US-0060837P.
PR 02-OCT-1997; 97US-0060838P.
PR 02-OCT-1997; 97US-0060839P.
PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
PI Lafleur DW, Shi Y;
XX
DR WPI; 1999-264022/22.
DR P-PSDB; AAY07861.
XX
PT New isolated human genes and the secreted polypeptides they encode.
XX
PS Claim 1a; Page 232-233; 368pp; English.
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552
XX
SQ Sequence 1089 BP; 367 A; 202 C; 240 G; 279 T; 0 U; 1 Other;

Query Match 40.5%; Score 15; DB 2; Length 1089;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GGAGAAGATCGCTCA 17
||||||||||||||||
Db 14 GGAGAAGATCGCTCA 28

RESULT 13
AAA28524
ID AAA28524 standard; cDNA; 1601 BP.
XX
AC AAA28524;
XX
DT 29-AUG-2000 (first entry)
XX
DE Human opioid growth factor receptor cDNA spliced version 1.
XX
KW OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1. .33
FT /*tag= a
FT CDS 34. .1419
FT /*tag= b
FT /product= "Opioid_growth_factor_receptor"
FT 3'UTR 1420. .1601
FT /*tag= c
XX
PN WO200026340-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-US025802.
XX
PR 03-NOV-1998; 98US-0106879P.
XX
PA (PENN-) PENN STATE RES FOUND.
XX
PI Zagon IS, Mclaughlin PJ, Verderame MF;
XX
DR WPI; 2000-365594/31.
DR P-PSDB; AAY92810.
XX
PT New cDNA encoding rat and human opioid growth factor receptors which
PT modulate cell growth, useful for treating cancer.
XX
PS Claim 1; Page 82-83; 91pp; English.
XX
CC Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC consistently yielded a single species of cDNA, while the 3' RACE revealed
CC extensive alternative splicing. The alternate splice forms were missing
CC the imperfect repeats or differed in the number of imperfect repeats. The
CC human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC OGFr proteins, nucleic acid molecules, antibodies, transformed cells and
CC expression vector are useful for detecting expression or levels of an
CC OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC cells in vitro. The antisense sequences and antibodies can be used to
CC promote growth of cells in vitro. Cell growth can be promoted by
CC interfering with the OGF ligand-receptor system, especially where a
CC subject suffers from a tissue wound. Treating cancer comprises enhancing
CC the function of the OGF ligand-receptor system in cancerous cells of a
CC patient or administering the OGFr nucleic acid to the patient
XX
SQ Sequence 1601 BP; 322 A; 485 C; 558 G; 236 T; 0 U; 0 Other;

Query Match 40.5%; Score 15; DB 3; Length 1601;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TGGAGAAGATCGCTC 16
|||||||||||||||
Db 1253 TGGAGAAGATCGCTC 1267

RESULT 14
ADD44906/c
ID ADD44906 standard; DNA; 1756 BP.
XX
AC ADD44906;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat gene Y15054, SEQ ID NO 10337.
XX
KW Rat; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX

```
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; Y15054.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC   specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1756 BP; 355 A; 543 C; 484 G; 374 T; 0 U; 0 Other;

  Query Match             40.5%;  Score 15;  DB 10;  Length 1756;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         13 GCTCAGTCTAGCCTG 27
              |||||||||||||||
Db        402 GCTCAGTCTAGCCTG 388


RESULT 15
AAA28526
ID   AAA28526 standard; cDNA; 2289 BP.
XX
AC   AAA28526;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Human opioid growth factor receptor cDNA spliced version 7.
XX
KW   OGFr; opioid growth factor receptor; growth inhibitor; proliferative;
KW   cytostatic; vulnerary. gene therapy; antagonist; chromosome 20q13.3; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .33
FT                   /*tag=  a
FT   CDS             34. .2007
FT                   /*tag=  b
FT                   /product= "Opioid_growth_factor_receptor"
FT   3'UTR           2008. .2289
FT                   /*tag=  c
XX
PN   WO200026340-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-US025802.
XX
PR   03-NOV-1998;   98US-0106879P.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Zagon IS,  Mclaughlin PJ,  Verderame MF;
XX
DR   WPI; 2000-365594/31.
DR   P-PSDB; AAY92812.
XX
PT   New cDNA encoding rat and human opioid growth factor receptors which
PT   modulate cell growth, useful for treating cancer.
XX
PS   Claim 1; Page 87-89; 91pp; English.
XX
CC   Primers generated from rat opioid growth factor receptor (OGFr) cDNA were
CC   used to clone a fragment of the human OGFr cDNA. The complete sequence of
CC   human OGFr was assembled with a combination of 3' and 5' RACE. 5' RACE
CC   consistently yielded a single species of cDNA, while the 3' RACE revealed
CC   extensive alternative splicing. The alternate splice forms were missing
CC   the imperfect repeats or differed in the number of imperfect repeats. The
CC   human OGFr gene chromosomal location was determined by FISH as 20q13.3.
CC   OGFr proteins, nucleic acid molecules, antibodies, transformed cells and
CC   expression vector are useful for detecting expression or levels of an
CC   OGFr in a tissue. OGFr nucleic acids can be used to inhibit growth of
CC   cells in vitro. The antisense sequences and antibodies can be used to
CC   promote growth of cells in vitro. Cell growth can be promoted by
CC   interfering with the OGF ligand-receptor system, especially where a
CC   subject suffers from a tissue wound. Treating cancer comprises enhancing
CC   the function of the OGF ligand-receptor system in cancerous cells of a
CC   patient or administering the OGFr nucleic acid to the patient
XX
SQ   Sequence 2289 BP; 470 A; 714 C; 809 G; 296 T; 0 U; 0 Other;

  Query Match             40.5%;  Score 15;  DB 3;  Length 2289;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TGGAGAAGATCGCTC 16
              |||||||||||||||
Db       1253 TGGAGAAGATCGCTC 1267


Search completed: December 22, 2004, 22:44:23
Job time : 136.838 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 22:08:48 ; Search time 29.8009 Seconds
(without alignments)
882.496 Million cell updates/sec

Title: US-09-898-616A-4
Perfect score: 37
Sequence: 1 atggagaagatcgctcagtctagcctgtgcaactaag 37

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		37	100.0	37	3	US-08-864-357F-9	Sequence 9, Appli
2	c	37	100.0	60	3	US-08-864-357F-10	Sequence 10, Appl
3		15	40.5	3715	4	US-09-234-245-1	Sequence 1, Appli
4	c	14	37.8	283	3	US-08-760-534A-5	Sequence 5, Appli
5	c	14	37.8	283	4	US-09-336-757-5	Sequence 5, Appli
6	c	14	37.8	341	4	US-09-513-999C-3565	Sequence 3565, Ap
7		14	37.8	422	4	US-09-621-976-12461	Sequence 12461, A
8		14	37.8	553	1	US-08-463-115-26	Sequence 26, Appl
9		14	37.8	553	1	US-08-465-388-26	Sequence 26, Appl
10		14	37.8	715	1	US-08-463-115-18	Sequence 18, Appl
11		14	37.8	715	1	US-08-465-388-18	Sequence 18, Appl
12	c	14	37.8	1374	4	US-09-252-991A-5999	Sequence 5999, Ap
13		14	37.8	1540	1	US-08-463-115-2	Sequence 2, Appli
14		14	37.8	1540	1	US-08-465-388-2	Sequence 2, Appli
15		14	37.8	1899	4	US-09-252-991A-6306	Sequence 6306, Ap
16	c	14	37.8	2118	4	US-09-799-451-298	Sequence 298, App
17	c	14	37.8	2196	1	US-08-472-217-3	Sequence 3, Appli
18	c	14	37.8	2196	3	US-08-760-534A-3	Sequence 3, Appli
19	c	14	37.8	2196	4	US-09-336-757-3	Sequence 3, Appli
20	c	14	37.8	2895	4	US-09-556-877-171	Sequence 171, App
21	c	14	37.8	2895	4	US-09-620-412C-171	Sequence 171, App
22	c	14	37.8	2895	4	US-09-598-419-171	Sequence 171, App
23	c	14	37.8	2934	4	US-09-556-877-183	Sequence 183, App
24	c	14	37.8	2934	4	US-09-620-412C-183	Sequence 183, App
25	c	14	37.8	2934	4	US-09-598-419-183	Sequence 183, App
26	c	14	37.8	40000	4	US-09-780-049-18	Sequence 18, Appl
27	c	14	37.8	72604	3	US-09-268-992-7	Sequence 7, Appli
28	c	14	37.8	72604	3	US-09-657-474-7	Sequence 7, Appli
29	c	14	37.8	1830121	4	US-09-557-884-1	Sequence 1, Appli
30	c	14	37.8	1830121	4	US-09-643-990A-1	Sequence 1, Appli
31	c	14	37.8	1830121	4	US-10-329-960-1	Sequence 1, Appli
32	c	14	37.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
33	c	14	37.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
34		13	35.1	271	4	US-09-313-294A-695	Sequence 695, App
35		13	35.1	289	3	US-08-651-155B-12	Sequence 12, Appl
36		13	35.1	289	4	US-09-194-036B-12	Sequence 12, Appl
37		13	35.1	297	4	US-09-252-991A-8030	Sequence 8030, Ap
38		13	35.1	301	4	US-09-270-767-2113	Sequence 2113, Ap
39		13	35.1	301	4	US-09-270-767-17395	Sequence 17395, A
40		13	35.1	342	4	US-10-101-464A-147	Sequence 147, App
41		13	35.1	348	3	US-08-844-059-3	Sequence 3, Appli
42		13	35.1	348	3	US-09-431-202-3	Sequence 3, Appli
43		13	35.1	369	4	US-09-270-767-14734	Sequence 14734, A
44		13	35.1	410	4	US-09-221-017B-273	Sequence 273, App
45		13	35.1	423	4	US-09-252-991A-11388	Sequence 11388, A

ALIGNMENTS

RESULT 1
US-08-864-357F-9
; Sequence 9, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 37
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-9

Query Match 100.0%; Score 37; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.2e-12;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37

RESULT 2
US-08-864-357F-10/c
; Sequence 10, Application US/08864357F
; Patent No. 6255281
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/08/864,357F
; CURRENT FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 60
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: primer sequence
US-08-864-357F-10

```
  Query Match             100.0%;  Score 37;  DB 3;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 3
US-09-234-245-1
; Sequence 1, Application US/09234245
; Patent No. 6509457
; GENERAL INFORMATION:
;  APPLICANT: Thomas Jefferson University
;  APPLICANT: Raphael Rubin
;  APPLICANT: Manorama Tewari
;  TITLE OF INVENTION: Compositions And Methods For Modulating
;  TITLE OF INVENTION: Apoptosis
;  FILE REFERENCE: TJU-002
;  CURRENT APPLICATION NUMBER: US/09/234,245
;  CURRENT FILING DATE: 1999-01-20
;  EARLIER APPLICATION NUMBER: US 60/071,878
;  EARLIER FILING DATE: 1998-01-20
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 3715
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-234-245-1

  Query Match             40.5%;  Score 15;  DB 4;  Length 3715;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          15 TCAGTCTAGCCTGTG 29
               |||||||||||||||
Db        1976 TCAGTCTAGCCTGTG 1990


RESULT 4
US-08-760-534A-5/c
; Sequence 5, Application US/08760534A
; Patent No. 6017727
;  GENERAL INFORMATION:
;    APPLICANT:  JALKANEN, MARKKU
;    APPLICANT:  JAAKKOLA, PANU
;    APPLICANT:  VIHINEN, TAPANI
;    TITLE OF INVENTION:  SYNDECAN ENHANCER ELEMENT AND SYNDECAN
;    TITLE OF INVENTION:  STIMULATION OF CELLULAR DIFFERENTIATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/760,534A
;      FILING DATE:  02-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,186
;      FILING DATE:  07-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI93/00514
;      FILING DATE:  01-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIMBALA, MICHELE A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1708.0050004/MAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-760-534A-5

  Query Match             37.8%;  Score 14;  DB 3;  Length 283;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          15 TCAGTCTAGCCTGT 28
               ||||||||||||||
Db         269 TCAGTCTAGCCTGT 256


RESULT 5
US-09-336-757-5/c
; Sequence 5, Application US/09336757
; Patent No. 6492344
;  GENERAL INFORMATION:
;    APPLICANT:  JALKANEN, MARKKU
;    APPLICANT:  JAAKKOLA, PANU
;    APPLICANT:  VIHINEN, TAPANI
;    TITLE OF INVENTION:  SYNDECAN ENHANCER ELEMENT AND SYNDECAN
;    TITLE OF INVENTION:  STIMULATION OF CELLULAR DIFFERENTIATION
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/336,757
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/760,534
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI93/00514
;      FILING DATE:  01-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIMBALA, MICHELE A.
;      REGISTRATION NUMBER:  33,851
;      REFERENCE/DOCKET NUMBER:  1708.0050004/MAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
US-09-336-757-5

  Query Match             37.8%;  Score 14;  DB 4;  Length 283;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 TCAGTCTAGCCTGT 28
              ||||||||||||||
Db        269 TCAGTCTAGCCTGT 256


RESULT 6
US-09-513-999C-3565/c
; Sequence 3565, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 3565
;   LENGTH: 341
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 22..339
US-09-513-999C-3565

  Query Match             37.8%;  Score 14;  DB 4;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         16 CAGTCTAGCCTGTG 29
              ||||||||||||||
Db        111 CAGTCTAGCCTGTG 98


RESULT 7
US-09-621-976-12461
; Sequence 12461, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 12461
;   LENGTH: 422
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-12461

  Query Match             37.8%;  Score 14;  DB 4;  Length 422;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         20 CTAGCCTGTGCAAC 33
              ||||||||||||||
Db        354 CTAGCCTGTGCAAC 367


RESULT 8
US-08-463-115-26
; Sequence 26, Application US/08463115
; Patent No. 5703221
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,115
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:             Four
;      APPLICATION NUMBER:  08/157,811
;      FILING DATE:  No. 5703221ember 23, 1993
;      APPLICATION NUMBER:  07/887,502
;      FILING DATE:  May 22, 1992
;      APPLICATION NUMBER:  07/704,814
;      FILING DATE:  May 23, 1991
;      APPLICATION NUMBER:  07/763,039
;      FILING DATE:  September 20, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  213/301
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:
US-08-463-115-26

  Query Match             37.8%;  Score 14;  DB 1;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAA 32
              ||||||||||||||
Db        310 TCTAGCCTGTGCAA 323


RESULT 9
US-08-465-388-26
; Sequence 26, Application US/08465388
; Patent No. 5753488
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,388
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:            Four
;      APPLICATION NUMBER:  08/157,811
;      FILING DATE:  No. 5753488ember 23, 1993
;      APPLICATION NUMBER:  07/887,502
;      FILING DATE:  May 22, 1992
;      APPLICATION NUMBER:  07/704,814
;      FILING DATE:  May 23, 1991
;      APPLICATION NUMBER:  07/763,039
;      FILING DATE:  September 20, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  213/300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:
US-08-465-388-26

  Query Match             37.8%;  Score 14;  DB 1;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAA 32
              ||||||||||||||
Db        310 TCTAGCCTGTGCAA 323


RESULT 10
US-08-463-115-18
; Sequence 18, Application US/08463115
; Patent No. 5703221
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,115
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:            Four
;      APPLICATION NUMBER:  08/157,811
;      FILING DATE:  No. 5703221ember 23, 1993
;      APPLICATION NUMBER:  07/887,502
;      FILING DATE:  May 22, 1992
;      APPLICATION NUMBER:  07/704,814
;      FILING DATE:  May 23, 1991
;      APPLICATION NUMBER:  07/763,039
;      FILING DATE:  September 20, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  213/301
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  715 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:
US-08-463-115-18

  Query Match             37.8%;  Score 14;  DB 1;  Length 715;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAA 32
              ||||||||||||||
Db        325 TCTAGCCTGTGCAA 338


RESULT 11
US-08-465-388-18
; Sequence 18, Application US/08465388
; Patent No. 5753488
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
```

```
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  FastSeq Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,388
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:              Four
;       APPLICATION NUMBER:  08/157,811
;       FILING DATE:  No. 5753488ember 23, 1993
;       APPLICATION NUMBER:  07/887,502
;       FILING DATE:  May 22, 1992
;       APPLICATION NUMBER:  07/704,814
;       FILING DATE:  May 23, 1991
;       APPLICATION NUMBER:  07/763,039
;       FILING DATE:  September 20, 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  213/300
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  715 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       OTHER INFORMATION:
US-08-465-388-18

  Query Match             37.8%;  Score 14;  DB 1;  Length 715;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          19 TCTAGCCTGTGCAA 32
               ||||||||||||||
Db         325 TCTAGCCTGTGCAA 338


RESULT 12
US-09-252-991A-5999/c
; Sequence 5999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5999
;   LENGTH: 1374
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5999

  Query Match             37.8%;  Score 14;  DB 4;  Length 1374;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           4 GAGAAGATCGCTCA 17
               ||||||||||||||
Db          78 GAGAAGATCGCTCA 65


RESULT 13
US-08-463-115-2
; Sequence 2, Application US/08463115
; Patent No. 5703221
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,115
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:              Four
;      APPLICATION NUMBER:  08/157,811
;      FILING DATE:  No. 5703221ember 23, 1993
;      APPLICATION NUMBER:  07/887,502
;      FILING DATE:  May 22, 1992
;      APPLICATION NUMBER:  07/704,814
;      FILING DATE:  May 23, 1991
;      APPLICATION NUMBER:  07/763,039
;      FILING DATE:  September 20, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  213/301
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-463-115-2

  Query Match             37.8%;  Score 14;  DB 1;  Length 1540;
  Best Local Similarity   100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          19 TCTAGCCTGTGCAA 32
               ||||||||||||||
Db         997 TCTAGCCTGTGCAA 1010


RESULT 14
US-08-465-388-2
; Sequence 2, Application US/08465388
; Patent No. 5753488
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM JOHN MARTIN
;    TITLE OF INVENTION:  ISOLATED STEALTH VIRUSES
;    TITLE OF INVENTION:  AND RELATED VACCINES
```

```
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,388
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:           Four
;      APPLICATION NUMBER:  08/157,811
;      FILING DATE:  No. 5753488ember 23, 1993
;      APPLICATION NUMBER:  07/887,502
;      FILING DATE:  May 22, 1992
;      APPLICATION NUMBER:  07/704,814
;      FILING DATE:  May 23, 1991
;      APPLICATION NUMBER:  07/763,039
;      FILING DATE:  September 20, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  213/300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-465-388-2

  Query Match            37.8%;  Score 14;  DB 1;  Length 1540;
  Best Local Similarity  100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAA 32
              ||||||||||||||
Db        997 TCTAGCCTGTGCAA 1010


RESULT 15
US-09-252-991A-6306
; Sequence 6306, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6306
;   LENGTH: 1899
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6306

  Query Match            37.8%;  Score 14;  DB 4;  Length 1899;
  Best Local Similarity  100.0%;  Pred. No. 43;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GAGAAGATCGCTCA 17
              ||||||||||||||
Db       1534 GAGAAGATCGCTCA 1547


Search completed: December 23, 2004, 01:33:53
Job time : 40.8009 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: December 22, 2004, 23:36:53 ; Search time 518.67 Seconds
(without alignments)
397.214 Million cell updates/sec

Title: US-09-898-616A-4
Perfect score: 37
Sequence: 1 atggagaagatcgctcagtctagcctgtgcaactaag 37

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4105333 seqs, 2784095677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8210666

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	100.0	37	9	US-09-861-688-9	Sequence 9, Appli
2	37	100.0	37	10	US-09-898-616A-4	Sequence 4, Appli
3	37	100.0	37	15	US-10-187-498A-4	Sequence 4, Appli
4	37	100.0	37	16	US-10-647-371-8	Sequence 8, Appli
c 5	37	100.0	60	9	US-09-861-688-10	Sequence 10, Appl
c 6	37	100.0	60	10	US-09-898-616A-5	Sequence 5, Appli
c 7	37	100.0	60	15	US-10-187-498A-5	Sequence 5, Appli
c 8	37	100.0	60	16	US-10-647-371-9	Sequence 9, Appli
9	16	43.2	32706	13	US-10-087-192-1750	Sequence 1750, Ap
c 10	16	43.2	78268	13	US-10-087-192-742	Sequence 742, App
c 11	15	40.5	261	18	US-10-674-124A-5867	Sequence 5867, Ap
c 12	15	40.5	389	16	US-10-424-599-75149	Sequence 75149, A
13	15	40.5	477	10	US-09-918-995-1312	Sequence 1312, Ap
c 14	15	40.5	574	13	US-10-027-632-131785	Sequence 131785,
c 15	15	40.5	574	13	US-10-027-632-131786	Sequence 131786,
c 16	15	40.5	574	15	US-10-027-632-131785	Sequence 131785,
c 17	15	40.5	574	15	US-10-027-632-131786	Sequence 131786,
c 18	15	40.5	679	16	US-10-424-599-82650	Sequence 82650, A
19	15	40.5	686	16	US-10-424-599-109965	Sequence 109965,
c 20	15	40.5	727	18	US-10-425-115-90184	Sequence 90184, A
c 21	15	40.5	1043	10	US-09-397-945-82	Sequence 82, Appl
c 22	15	40.5	1043	16	US-10-653-595-82	Sequence 82, Appl
23	15	40.5	1089	15	US-10-195-730-20	Sequence 20, Appl
24	15	40.5	1089	17	US-10-799-747-20	Sequence 20, Appl
25	15	40.5	1224	17	US-10-437-963-94846	Sequence 94846, A
26	15	40.5	1436	17	US-10-437-963-46992	Sequence 46992, A
c 27	15	40.5	1946	17	US-10-767-701-14898	Sequence 14898, A
28	15	40.5	2201	10	US-09-791-254-4	Sequence 4, Appli
29	15	40.5	2814	17	US-10-437-963-94851	Sequence 94851, A
30	15	40.5	2823	17	US-10-437-963-94806	Sequence 94806, A
31	15	40.5	2835	17	US-10-437-963-94776	Sequence 94776, A
32	15	40.5	2889	17	US-10-437-963-94807	Sequence 94807, A
33	15	40.5	2952	17	US-10-437-963-94766	Sequence 94766, A
34	15	40.5	3366	17	US-10-437-963-47007	Sequence 47007, A
35	15	40.5	3396	17	US-10-437-963-94769	Sequence 94769, A
36	15	40.5	3414	17	US-10-437-963-94802	Sequence 94802, A
37	15	40.5	3501	17	US-10-437-963-94774	Sequence 94774, A
38	15	40.5	3558	17	US-10-437-963-47002	Sequence 47002, A
39	15	40.5	3561	17	US-10-437-963-94836	Sequence 94836, A
40	15	40.5	3591	17	US-10-437-963-94843	Sequence 94843, A
41	15	40.5	3618	17	US-10-437-963-94841	Sequence 94841, A
42	15	40.5	3677	17	US-10-437-963-94855	Sequence 94855, A
c 43	15	40.5	3687	17	US-10-437-963-47011	Sequence 47011, A
44	15	40.5	3711	17	US-10-437-963-94812	Sequence 94812, A
45	15	40.5	4065	10	US-09-791-254-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-861-688-9
; Sequence 9, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Claragen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 08/864,357
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-9

Query Match 100.0%; Score 37; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
||||||||||||||||||||||||||||||||||||||
Db 1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37

RESULT 2
US-09-898-616A-4

```
; Sequence 4, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 37
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-4

  Query Match           100.0%;  Score 37;  DB 10;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37


RESULT 3
US-10-187-498A-4
; Sequence 4, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 37
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-4

  Query Match           100.0%;  Score 37;  DB 15;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37


RESULT 4
US-10-647-371-8
; Sequence 8, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 37
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-8

  Query Match           100.0%;  Score 37;  DB 16;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37


RESULT 5
US-09-861-688-10/c
; Sequence 10, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of
;  TITLE OF INVENTION:  Inflammatory and
;  TITLE OF INVENTION:  Fibrotic Conditions
;  FILE REFERENCE: 116142/2
;  CURRENT APPLICATION NUMBER: US/09/861,688
;  CURRENT FILING DATE:  2001-05-21
;  PRIOR APPLICATION NUMBER: 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: primer sequence
US-09-861-688-10

  Query Match           100.0%;  Score 37;  DB 9;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 6
US-09-898-616A-5/c
; Sequence 5, Application US/09898616A
; Publication No. US20030109429A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00170
;  CURRENT APPLICATION NUMBER: US/09/898,616A
;  CURRENT FILING DATE:  2002-10-15
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-09-898-616A-5

  Query Match             100.0%;  Score 37;  DB 10;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 7
US-10-187-498A-5/c
; Sequence 5, Application US/10187498A
; Publication No. US20030207795A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen Inc.
;  APPLICANT:  Pilon, Aprile L
;  APPLICANT:  Welch, Richard W
;  TITLE OF INVENTION: Method for the Production of Purified rhUG for the Treatment of
;  TITLE OF INVENTION:  Imflammatory and Fibrotic Conditions
;  FILE REFERENCE: 116142/00260
;  CURRENT APPLICATION NUMBER: US/10/187,498A
;  CURRENT FILING DATE:  2001-07-02
;  PRIOR APPLICATION NUMBER: US 08/864,357
;  PRIOR FILING DATE: 1997-05-28
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Original amino acid sequence obtained from cDNA.  Current sumitte
;   OTHER INFORMATION: d sequence maximized for expression in E. coli.
US-10-187-498A-5

  Query Match             100.0%;  Score 37;  DB 15;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 8
US-10-647-371-9/c
; Sequence 9, Application US/10647371
; Publication No. US20040047857A1
; GENERAL INFORMATION:
;  APPLICANT: Claragen, Inc. & NIH
;  TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammatory
;  TITLE OF INVENTION:  and Fibrotic Conditions
;  FILE REFERENCE: 116142-85
;  CURRENT APPLICATION NUMBER: US/10/647,371
;  CURRENT FILING DATE:  2003-08-25
;  PRIOR APPLICATION NUMBER: 09/549,926
;  PRIOR FILING DATE: 2000-04-14
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  primer sequence
US-10-647-371-9

  Query Match             100.0%;  Score 37;  DB 16;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-12;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 37
              |||||||||||||||||||||||||||||||||||||
Db         37 ATGGAGAAGATCGCTCAGTCTAGCCTGTGCAACTAAG 1


RESULT 9
US-10-087-192-1750
; Sequence 1750, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1750
;   LENGTH: 32706
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-087-192-1750

  Query Match              43.2%;  Score 16;  DB 13;  Length 32706;
  Best Local Similarity   100.0%;  Pred. No. 7.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAACT 34
              ||||||||||||||||
Db       7293 TCTAGCCTGTGCAACT 7308


RESULT 10
US-10-087-192-742/c
; Sequence 742, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
```

```
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 742
;   LENGTH: 78268
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-087-192-742

  Query Match             43.2%;  Score 16;  DB 13;  Length 78268;
  Best Local Similarity   100.0%;  Pred. No. 7.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         13 GCTCAGTCTAGCCTGT 28
               ||||||||||||||||
Db      29223 GCTCAGTCTAGCCTGT 29208


RESULT 11
US-10-674-124A-5867/c
; Sequence 5867, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 5867
;   LENGTH: 261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: AC072028.5_149909
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 3
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 147898939
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 130675
US-10-674-124A-5867

  Query Match             40.5%;  Score 15;  DB 18;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 TCTAGCCTGTGCAAC 33
               |||||||||||||||
Db        240 TCTAGCCTGTGCAAC 226


RESULT 12
US-10-424-599-75149/c
; Sequence 75149, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 75149
;   LENGTH: 389
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_38873C.1
US-10-424-599-75149

  Query Match             40.5%;  Score 15;  DB 16;  Length 389;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 TCAGTCTAGCCTGTG 29
               |||||||||||||||
Db         40 TCAGTCTAGCCTGTG 26


RESULT 13
US-09-918-995-1312
; Sequence 1312, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1312
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(477)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-1312

  Query Match             40.5%;  Score 15;  DB 10;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TGGAGAAGATCGCTC 16
               |||||||||||||||
Db        305 TGGAGAAGATCGCTC 319


RESULT 14
US-10-027-632-131785/c
; Sequence 131785, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 131785
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-131785

Query Match 40.5%; Score 15; DB 13; Length 574;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 GTCTAGCCTGTGCAA 32
||||||||||||||||
Db 40 GTCTAGCCTGTGCAA 26

RESULT 15
US-10-027-632-131786/c
; Sequence 131786, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 131786
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-131786

Query Match 40.5%; Score 15; DB 13; Length 574;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 GTCTAGCCTGTGCAA 32
||||||||||||||||
Db 40 GTCTAGCCTGTGCAA 26

Search completed: December 23, 2004, 05:19:30
Job time : 520.67 secs